Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
HUMAN RESOURCES RESEARCH CENTER
LACKLAND AIR FORCE BASE
SAN ANTONIO, TEXAS

CONSULTING EDITORS

Judson S. Brown, *State University of Iowa*
Frank A. Geldard, *University of Virginia*
James J. Gibson, *Cornell University*
Clarence H. Graham, *Columbia University*
David A. Grant, *University of Wisconsin*
Lloyd G. Humphreys, *Human Resources Research Center*
Donald B. Lindsley, *University of California, Los Angeles*
Neal E. Miller, *Yale University*
Edwin B. Newman, *Harvard University*
Benton J. Underwood, *Northwestern University*
Delos D. Wickens, *Ohio State University*

Lorraine Bouthilet, Managing Editor

THE JOHN CRERAR
1952
LIBRARY

CONTENTS

A Study of the Transposition Gradient: D. EHRENFREUND 81
Rote Learning as a Function of Distribution of Practice and the Complexity of the Situation: D. A. RILEY 88
The Decline of Pitch Discrimination With Time: J. D. HARRIS 96
Effects of Failure Stress upon Skilled Performance: R. S. LAZARUS AND C. W. ERIKSEN .. 100
Response Strength in a Modified Thorndikian Multiple-Choice Situation as a Function of Varying Proportions of Reinforcement: A. E. GOSS AND E. J. RABAIOLI 106
Motivation in Learning: XI. An Analysis of Electric Shock for Correct Responses into its Avoidance and Accelerating Components: K. F. MUENZINGER, W. O. BROWN, W. J. CROW, AND R. F. POWLOSKI 115
The Effect of Degree of Shading Contrast in Ink Blots on Verbal Response: J. C. BALLOCH 120
Rate of Manipulative Learning as a Function of Goal-Setting Techniques: G. C. HELMSTADTER AND D. S. ELLIS 125
Reactive Inhibition as a Factor in Maze Learning: III. Effects in the Human Stylus Maze: M. E. THOMPSON 130
Latent Learning in a Water Maze: A. W. BENDIG 134
Studies in Short-Duration Auditory Fatigue: II. Recovery Time: A. I. RAWNSLEY AND J. D. HARRIS 138
Studies of Distributed Practice: V. Learning and Retention of Concepts: L. OSEAS AND B. J. UNDERWOOD 143
Sensitivity to Apparent Movement in Depth as a Function of Stimulus Dimensionality: W. M. SMITH 149
The Role of Repetition Rate and Inter-Stimulus Interval in Context Effects: J. M. DOUGHTY 156

American Psychological Association

Vol. 43 No. 2 February 1952

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc. The subscription rate per volume is $7.00, or $14.00 annually. Single copies are $1.25. Subscriptions, orders, and business communications should be addressed to the American Psychological Association, Inc., Prince and Lemon Sts., Lancaster, Pa., or 1515 Massachusetts Avenue, N.W., Washington 5, D. C.

This JOURNAL publishes articles which are intended to contribute toward the development of psychology as an experimental science. Experimental work with normal human subjects is favored over work with abnormal or animal subjects. Studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental psychology or for behavior theory. Normally, articles of a length exceeding 15 printed pages cannot be accepted. However, a single experimental study or an integrated series of studies accomplished simultaneously must be presented in a single article, rather than in a series of articles.

Articles are published in the order of their receipt, except in rare circumstances. Authors who request them are supplied with 50 free reprints without covers. However, under the new APA reprint policy no free reprints will be supplied to authors submitting manuscripts after January 1, 1952. All of the cost of the author's alterations in galley proof is charged to the author, and one-half of the cost of tables, cuts, and formulas.

Immediate publication can be given to articles whose authors assume the full cost of publication. Such articles must undergo the usual review by the Consulting Editors of the JOURNAL. No free reprints are given in the case of these articles.

Address all articles submitted for regular publication or for immediate publication to the editor, Arthur W. Melton, Human Resources Research Center, Lackland Air Force Base, San Antonio, Texas.

The JOURNAL adheres to the conventions concerning bibliographies, citations, preparation of tables and cuts, etc., as described by J. E. Anderson and W. L. Valentine, The preparation of articles for publication in the journals of the American Psychological Association, *Psychol. Bull.*, 1944, 41, 345–376. Special attention is called to the style of numbers and abbreviations (p. 353). The term 'experimenter,' with its various forms, singular, plural, and possessive singular and plural, is abbreviated *E*, *Es*, *E*'s, and *Es*'; analogous abbreviations are used for the terms 'subject' (*S*) and 'observer' (*O*). All manuscripts must be typed double-spaced throughout, including tables, footnotes, references, and titles for figures. Tables, figures, footnotes, references, and titles for figures must be placed on separate pages at the end of the manuscript.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879.

Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

Journal of
Experimental Psychology

VOL. 43, NO. 2 FEBRUARY, 1952

A STUDY OF THE TRANSPOSITION GRADIENT [1]

DAVID EHRENFREUND
The State College of Washington

Despite the mass of data collected on the white rat, a recent study by Kendler (12) represents the only report to date in which the transposition phenomenon has been systematically investigated with this *S*. This paucity of data perhaps accounts for Munn's omission of the theoretical implications of transposition in his chapter on systematic psychology. In discussing the data, however, he recognizes its theoretical import and accepts unquestioningly the gestalt notion that the response, initially at least, is to a relationship (16, pp. 131, 143, 178). Other recent publications (1, 5, 11, 15) not restricted to rat data nonetheless arrive at the same conclusion despite evidence that should lead at least to a questioning of the adequacy of either the traditional or neo-gestalt viewpoint (3, 10, 13, 14, 18, 19, 20). Hebb (5), however, takes the position that although the response is to a relationship, this does not support the gestalt position.[2]

The above references (1, 5, 11, 15) have restricted themselves almost exclusively to the transposition experiment which limits its test to but one step removed from the training pair, e.g., *Ss* are first trained on medium gray positive versus light gray negative and then tested on a medium gray opposed to a dark gray. Under certain conditions, the responses on such a test are to the dark gray. Since the *Ss* were originally trained on the "darker than" and respond likewise in the test situation one might assume that the original discrimination was learned on the basis of responding to a relationship or configuration rather than to the absolute properties of each cue. Such assumptions, however, are neither logically demanded nor necessarily the most appropriate ones.

In their well-developed gestalt theory, Gulliksen and Wolfle discuss the illegitimacy of the relational assumption:

> *Transfer* on a relational basis is . . . interpreted as evidence that *learning* was on a relational basis. It has now been demonstrated that relative transfer can be predicted from a theory based on a directional

[1] This investigation was supported in part by The State College of Washington Research Fund. The writer wishes to thank Mr. Garry Margolius, research assistant, for collecting part of the data.

[2] Others such as Hilgard (6) confuse the issue of transposition with the related one of continuity by discussing data relevant to the latter to evaluate the former. This technique is all the more surprising in view of the more adequate treatment contained in the earlier Hilgard and Marquis publication (7).

> definition of the response. Since this prediction is possible, it is not legitimate to use the fact of relational transfer to support the assumption that the animal's original learning was relational in nature (4, p. 240).

Their formulation, however, does not account for instances of either choice scores or reversals in the test (transfer) situation.

A theory more capable in this respect is that proposed by Spence (16, 17, 18, 19).

> His theory, unlike the various Gestalt interpretations, does not assume that the animal is responding to some abstract relational property, or to a primitive structure-process. He holds, on the contrary, that the positive member in the stimulus-series acquires excitatory properties as a result of training to the response of approaching it and, furthermore, that there is a gradient of generalization of this excitation to other members of the stimulus-series, including the negative stimulus-object. With failure of reenforcement of response to the latter, however, an inhibitory tendency develops which, in turn, is assumed to generalize to adjoining members of the stimulus-series (19, p. 223).

The theoretical picture presented by Spence has successfully predicted not only the success or failure of transposition when tests are conducted on logarithmically equal whole-step stimulus pairs but also generates testable hypotheses concerning probability of transposition when the tests are not so restricted. In other words, the theory does not demand that the difference between the members of the test pair be equal to the difference employed in the original discrimination. Nor are the tests necessarily limited to such pairs as are whole-step units removed from the training pair. Employing a stimulus series that varies in a single dimension, the theory should predict the relative amount of transposition on any pair after training on any other pair. Theoretical confirmation from tests employing partial steps has been reported by Gulliksen (3). The prediction that transposition tends to decrease as the difference between the training and test pairs increases in whole-step units has also been confirmed (3, 12, 13, 14, 18, 19). However, the nature of the function combining these two situations has received little experimental attention.

In the present investigation, these two situations were combined by first training a large group of *S*s on a brightness discrimination in which the difference between the members was fairly large. Upon reaching the criterion of learning each *S* was tested on one of four new pairs of stimuli differing along a brightness dimension at various partial steps distant from the original.

Method

Subjects.—The *S*s were 80 albino male rats from the colony maintained by the psychology department of the State College of Washington.

Apparatus.—The apparatus was a simple T-maze design, the stem of which was walled and covered with hardware cloth. The arms were open as in an elevated maze. The floor consisted of a grid of ½-in. wire spaced ½ in. apart. Each arm, from the center of the choice point, could accommodate a 4 × 10-in. stimulus card. At the end of each runway, a glass food cup (furniture coaster) could be fastened to a two-pronged wire holder in such a way as to be invisible from the choice point. A T-shaped aluminum door at the choice point prevented both retracing and correction. The sole illumination was provided by a 20-w. fluorescent lamp shielded by two thicknesses of transluscent tracing paper. The stimuli for the brightness continuum consisted of eight 4 × 10-in. sheets of 060-gauge cellulose acetate plastic. These were purchased from the Container Corporation of America and are identical to the brightness series employed in the second edition of the Color Harmony Manual.[2] The main advantages of this material lie in its rigidity, homogene-

[2] Acknowledgment is extended to the helpful cooperation of Mr. Walter C. Granville and Mr. Carl E. Foss, editors of the Color Harmony Manual, for making available the stimuli employed in the present investigation.

TABLE 1

TRAINING PAIRS AND VARIOUS TEST PAIRS (A TO D) USED IN BOTH EXPERIMENTS

Stimulus Pairs Exp. I	Stimulus Cards Diffuse Reflection Factors	Ratio	Stimulus Pairs Exp. II
Training pair	(4)* .22 (1) .91	4.1:1	D
A	(5) .13 (2) .57	4.4:1	C
B	(6) .09 (3) .34	3.8:1	B
C	(7) .05 (4) .22	4.4:1	A
D	(8) .035 (5) .13	3.7:1	Training pair

* These numbers refer to stimulus cards.

ous matt surface, and ability to withstand repeated use and washing. The stimuli were calibrated by the Container Corporation as having the following diffuse reflection factors on their matt (painted) side: .91, .57, .34, .22, .13, .09, .05, .035. In the present discussion these stimuli will be referred to simply by number from 1 (white) to 8 (black).

Procedure.—After *Ss* were tamed and had experienced at least a week of being fed an 8-gm. daily ration of laboratory chow, they were introduced to the apparatus. On the first two days, each *S* was allowed to consume ten pellets of calf manna on each arm of the maze. During this time, the door was closed. On the next two days, each *S* was allowed ten runs from the starting box to the food cup where it consumed one pellet of food. These runs were forced left and right in a balanced random order. Learning of the original discrimination was then instituted. The appropriate stimulus cards were introduced for the first time and each *S* was given ten non-correction trials per day until the criterion of 18 out of 20 errorless trials, with the last ten errorless, was attained. During the training series, the food cup was placed only on the rewarded side. In Exp. I, *Ss* were trained on Stimulus 1 (−) vs. Stimulus 4 (+). In Exp. II, they were trained on Stimulus 8 (−) vs. Stimulus 5 (+). Each *S* was then tested for two days (20 trials) on a new pair A, B, C, or D as indicated in Table 1.

On the test trials, all responses were rewarded. Thus, in Exp. I, *Ss* were trained to the darker stimulus and transposed down the scale. In Exp. II, *Ss* were trained to the lighter stimulus and transposed up the scale. All transposition (test) scores are in terms of whether *S* responded in a manner consistent with the direction employed in training, i.e., to the darker or to the lighter. Ten *Ss* were tested on each condition of each experiment.

RESULTS

Trial and error scores for the original discrimination were fairly low. In Exp. I, in which *Ss* were trained to discriminate Stimulus 1 (non-rewarded) from Stimulus 4 (rewarded), the mean trials and errors were 86.9 and 42.6. In Exp. II (Stimulus 8 non-rewarded, Stimulus 5 rewarded), the mean trials and errors were 75.5, and 34.2.

The test trials for each test pair in both experiments were subdivided into successive five-trial blocks. The mean transposition scores obtained on these blocks of trials, as well as the total for all 20 trials, are presented in Table 2. The trends revealed in this table can be analyzed in terms of both columns and rows. Inspection of the rows, i.e., the subdivided test scores for any one test pair, reveals no consistent trend that will describe all of the tests. The scores for Test Pair A of Exp. I rise steadily from 86 to 94. On the other hand, the scores for Test Pair D of the same experiment fall

TABLE 2

PER CENT OF TRANSPOSITION BY FIVE-TRIAL BLOCKS

Tests	Trials				
	1–5	6–10	11–15	16–20	1–20
Exp. I Ss trained on Stimulus 1 (−) vs. Stimulus 4 (+)					
A (2–5)	86	88	92	94	90
B (3–6)	82	86	78	74	80
C (4–7)	48	42	46	40	44
D (5–8)	66	58	54	54	58
Exp. II Ss trained on Stimulus 8 (−) vs. Stimulus 5 (+)					
A (7–4)	84	84	84	80	83
B (6–3)	78	90	92	86	86.5
C (5–2)	28	26	20	28	25.5
D (4–1)	40	34	44	50	42

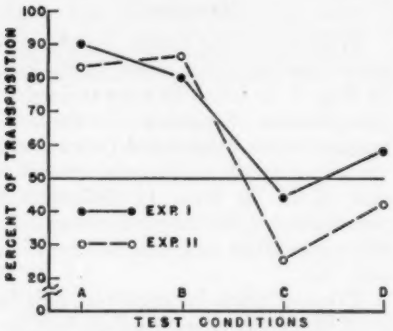

FIG. 1. Transposition gradients for *Ss* transposed down the brightness scale (Exp. I), and for *Ss* transposed up the brightness scale (Exp. II)

steadily from 66 to 54. Scores for other test pairs show various fluctuations. Apparently any block of trials is as appropriate as any other as a measure of transposition. The various columns within each experiment, however, all reveal similar tendencies.

The per cents of transposition, based on all 20 trials, are presented in Fig. 1. It will be recalled that in Exp. I, *Ss* were trained with the darker positive and tested on still darker pairs. Transposition is high on the near test and falls in an S-shaped fashion to a value below the 50% point and then rises above the 50% point. In Exp. II, *Ss* were trained on the lighter and tested on still lighter pairs. The data for Exp. II show the same general trends just described for Exp. I. A similar S-shaped tendency is evident except that the scores fall farther below the 50% point on Test Pair C (significant at the 1% level), and Test Pair D is still below 50%.

DISCUSSION

A recent study by T. S. Kendler (12) posed some interesting problems both in respect to the nature of the transposition function and the adequacy of Spence's theory. The nature of the function was difficult to determine in the Kendler study due to the fact that on any one test pair, the later test trials produced consistently higher scores than the earlier test trials. Therefore, the data were plotted using three response measures: Test Trials 1–5, 6–10, and 1–10 (12, Fig. 3 and 4). Inspection of these curves reveals that not only are they different, but no matter which curve is employed as the more appropriate function, it only partially corroborates the expectations from Spence's theory. Thus, although the amount of transposition decreased as a function of the difference between training and test stimuli, the test scores did not reach or go below the chance level as theoretically expected. According to the analysis made above for the present set of data, no such consistent trends between the subdivided test scores for the same test stimuli were discovered. This finding, coupled with the fact that the transposition function is essentially similar no matter which block of test trials is employed, justifies acceptance of the curves in Fig. 1 as at least close approximations to the empirical law relating amount of transposition to the distance between training and test stimuli along a brightness dimension.

Although the Kendler study was similar in many respects to the present one, she used direct lighting, and several complications attendant upon the use of direct rather than reflected light could easily account for the differences between the findings of her study and the present one. Some, such as the problem of dark adaptation and the changing efficiency of the rat's eye as trials continue, were discussed by Kendler. The possible necessity for positing "other factors in the present [Kendler] training or test situations in addition to those posited by Spence to

determine the final results" (12, p. 561) was suggested in a footnote but was not elaborated. These "other factors," which seem to the present author to be pertinent and peculiar to the direct illumination situation as employed by Kendler, include:

1. The change in extra-maze cues when *S* is shifted from the training to the test pair of stimuli. The use of direct illumination in an otherwise non-illuminated room results in a change in room cues from few to many or vice versa depending on the direction of training and testing. This change would, of course, be greatest for those groups tested on the farthest pair and least for those tested on the nearest pair.

2. Employing stimulus cues having drive value. The level of illumination employed on Test Pair 5 for Kendler's Cond. 1, or for training in Cond. 2, was of a magnitude comparable to that employed in other studies as noxious stimuli (21).

Thus, although the use of a situation employing direct illumination may not be the most appropriate one with which to test the specific assumptions posited by Spence to account for transposition, the Kendler study suggests a degree of complexity perhaps not heretofore appreciated and which merits further study, particularly in relation to attempts at more elaborate theoretical formulations. Furthermore, as Kendler points out, her results can be described within the framework of S-R reinforcement theory by employing, in addition to Spence's postulates, Hull's concepts of patterning of stimulus compounds (8) and his recently formulated stimulus dynamism (9).

Turning now to the implications of the present set of data for Spence's theory, we find a close approximation between prediction and results. Particular note should be made of the fact that the low point on each curve occurred on Test Pair C which is only one step removed from the training pair. In Exp. I, Test Pair C was composed of Stimulus 4, the positive stimulus employed in the training situation, and a new Stimulus 7. Instead of responding to Stimulus 7, as predicted by some form of relational theory, 56% of the responses were to the absolute stimulus, i.e., Stimulus 4. One might say that since the response was about chance, neither theoretical viewpoint was supported. According to CR theory, however, the amount of transposition will reach a chance level at some point, and under certain conditions will fail to occur, i.e., the response will be to the absolute stimulus (19, 20). Apparently this is what happened in Exp. II, where Test Pair C was composed of Stimulus 5, the reinforced stimulus of the training situation, and a new Stimulus 2. In this instance, 75% of the responses were to the absolute stimulus. Furthermore, nine of the ten *S*s responded better than 50% and three *S*s responded 100% to the absolute stimulus.

Although both transposition curves are in line with the general prediction from the theory, the difference between them may be important in terms of certain theoretical refinements. The generalization gradients of excitation and inhibition posited by Spence to account for transposition are symmetrical for both directions of the stimulus continuum (20). A recent study by Grice (2) indicates that for size, at least, such may not be the case. He found that the generalization gradient to the "larger" end of the continuum was of the same general shape as posited by Spence. However, toward the "smaller" end of the continuum, the gradient was discovered to be an exponential type not unlike that posited by Hull (8). There is also the question whether the generalization gradients for inhibition may be asymmetrical. Spence has suggested that ". . . the shape of the generalization curve [of inhibition] in the direction of larger stimulus-areas differs from that of the smaller" (19, p. 229).

If asymmetrical generalization gradients may be expected for the brightness dimension, as well as the size

dimension, then the difference between the two transposition gradients of the present study conforms to that expectation. In Exp. II there was a greater failure of transposition than in Exp. I. In fact, these data are almost exactly as predicted:

> That is, according to this schema, the per cent of responses consistent with the direction of the original training should at first decrease until they drop to well below the 50 per cent level (i.e., lead to a reversal of the originally learned preference) and subsequently rise and level off at a response value, presumably 50 per cent, where the difference between the theoretical effective excitatory strengths becomes zero (12, pp. 560–561).

Thus, it appears, from the present set of data, that Spence's theoretical generalization gradients are more appropriate for predicting transposition up the brightness scale than down. In Kendler's (**12**) study, the form of the curve was also more in line with the theory when *S*s were transposed up the brightness scale rather than down.

Summary

Two transposition experiments employing reflected light were conducted. In Exp. I, 40 albino rats were trained to respond to the darker of two bright stimuli and tested at various partial steps down the brightness scale. In Exp. II, 40 albino rats were trained to respond to the brighter of two dark stimuli and tested at various partial steps up the brightness scale. In each experiment, four tests were carried out at points increasingly distant from the training pair. Both transposition curves confirmed the general prediction from Spence's theory to the effect that transposition is less perfect as the difference between training and test pairs is increased.

The low point on each transposition gradient, which was either chance (Exp. I) or below chance (Exp. II), occurred on Test C which was only one step removed from the training pair. Thus, the data from Exp. II are more in line with the theory, being almost exactly as predicted. The differences between the two sets of data were discussed in terms of the possible need for positing asymmetrical generalization gradients.

(Manuscript received April 21, 1951)

References

1. Dashiell, J. F. *Fundamentals of general psychology.* Cambridge, Mass.: Houghton Mifflin, 1949.
2. Grice, G. R., & Saltz, E. The generalization of an instrumental response to stimuli varying in the size dimension. *J. exp. Psychol.*, 1950, **40**, 702–708.
3. Gulliksen, H. Studies of transfer of response: I. Relative versus absolute factors in the discrimination of size by the white rat. *J. genet. Psychol.*, 1932, **40**, 37–51.
4. Gulliksen, H., & Wolfle, D. L. A theory of learning and transfer: II. *Psychometrika*, 1938, **3**, 225–251.
5. Hebb, D. O. *The organization of behavior.* New York: John Wiley, 1949.
6. Hilgard, E. R. *Theories of learning.* New York: Appleton-Century-Crofts, 1948.
7. Hilgard, E. R., & Marquis, D. G. *Conditioning and learning.* New York: D. Appleton-Century, 1940.
8. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
9. Hull, C. L. Stimulus intensity dynamism (V) and stimulus generalization. *Psychol. Rev.*, 1949, **56**, 67–76.
10. Jones, F. N. The stepwise phenomenon in rats. *J. comp. Psychol.*, 1939, **27**, 39–44.
11. Keller, F. S., & Schoenfeld, W. N. *Principles of psychology.* New York: Appleton-Century-Crofts, 1950.
12. Kendler, T. S. An experimental investigation of transposition as a function of the difference between training and test stimuli. *J. exp. Psychol.*, 1950, **40**, 552–562.
13. Klüver, H. *Behavior mechanisms in monkeys.* Chicago: Univ. Chicago Press, 1933.
14. Kuenne, M. K. Experimental investiga-

tion of the relation of language to transposition behavior in young children. *J. exp. Psychol.*, 1946, **36**, 471–490.

15. Munn, N. L. *Psychology.* Cambridge, Mass.: Houghton Mifflin, 1946.
16. Munn, N. L. *Handbook of psychological research on the rat.* Cambridge, Mass.: Houghton Mifflin, 1950.
17. Spence, K. W. The nature of discrimination learning in animals. *Psychol. Rev.*, 1936, **43**, 427–449.
18. Spence, K. W. The differential response in animals to stimuli varying within a single dimension. *Psychol. Rev.*, 1937, **44**, 430–444.
19. Spence, K. W. The failure of transposition in size discrimination of chimpanzees. *Amer. J. Psychol.*, 1941, **54**, 223–229.
20. Spence, K. W. The basis of solution by chimpanzees of the intermediate size problem. *J. exp. Psychol.*, 1942, **31**, 257–271.
21. Zeaman, D., & House, B. J. Response latency at zero drive after varying numbers of reinforcements. *J. exp. Psychol.*, 1950, **40**, 570–583.

ROTE LEARNING AS A FUNCTION OF DISTRIBUTION OF PRACTICE AND THE COMPLEXITY OF THE SITUATION [1]

DONALD A. RILEY
University of California

The purpose of this study is to investigate the relationship between distribution of practice and the complexity of the situation in a verbal rote learning task, where complexity is defined in terms of the number of possible alternative responses which may be made to each stimulus.

Although Hovland (**10, 11**) has presented evidence that distribution of practice is superior to massed practice in serial and paired-associate learning situations in which the distribution intervals are of brief duration, investigators using conditions of greater complexity have found continuous practice superior to rest intervals of several hours between trials. Cook investigated the effects of distribution of practice in mazes of varying complexity, and found that in all situations an interval of 20 to 30 sec. was superior to a 4-hr. interval between trials, the relative superiority of massed practice being greater for the situations with the greater number of alternatives at each choice point.

Ericksen (**6**) compared the relative efficiency of continuous practice versus a brief practice period every other day in the learning of a multiple U-maze and a McGeoch puzzle box. On the basis of studies (**3, 5, 9, 22**) which had found that in rats, massing of trials brings about an increase in the variability of behavior in a T-maze, when two trials a day are given, he predicted that the maze problem should profit by distribution because it is a problem which, in the initial stages of the learning process, maximizes the fixative aspects of the process, while the puzzle problem should profit most by massing, because it is a problem which maximizes the variable aspects of the learning process in the initial stages of the problem. His prediction was confirmed, although no evidence was presented to show that variability of response was increased as a result of massing. That variability of response should increase with massing would seem to be a necessary consequence of the original assumption. Another prediction which would follow from the hypothesis that massing increases variability is that in simple situations, learning in the early trials would profit most by distributed practice; whereas, in the more complex situations, learning in the early trials would profit most by massed practice. This prediction would follow because in more complex situations increasing the variability of the response would increase the probability of discovering the correct response, whereas in simpler situations increasing the variability of response would increase the probability of giving a response other than the correct response after the correct response had been given. Distribution of practice should become more beneficial in the later stages of learning, where variability would interfere with the final fixation of the correct responses. Cook (**1, 2**) found

[1] This paper is a portion of a dissertation submitted to the graduate school of The Ohio State University in partial fulfillment of the requirements for the Ph.D. degree. The author wishes to express his appreciation to Drs. A. W. Melton, D. D. Wickens, and R. J. Wherry for their advice and assistance.

massing superior only in the early trials in puzzle solving, but in the maze studies he found massing superior throughout the learning process. Other earlier investigators (**14, 17**) found early distribution superior in simple situations with both humans and rats as *Ss*.

In consideration of the wide differences in *Ss*, materials to be learned, and methods of training, which are to be found in the various studies bearing on the problem of distribution of practice, it is not surprising that discrepant results have been obtained. The aim of this study is to hold the distribution intervals similar to those used in the Hovland studies and to make the simple condition comparable to the conventional paired-associate learning situation. Complexity is to be defined as the extent to which the discovery process is involved in the learning process, i.e., the greater the number of alternatives presented to the individual at each choice situation, the greater the complexity of the situation (**18**). This dimension is varied in a stepwise fashion.

The experiment being reported deals with the following questions: (a) What is the relationship between distribution of practice and the complexity of the situation as measured by learning and retention scores? (b) Is there evidence for greater variability of response in the massed conditions, and if so, what effect does this greater variability have on the efficiency of learning in the early and later parts of the learning process?

Procedure and Materials

Terminology.—The terms used to identify the various conditions will aid in describing the procedure. There were six groups of *Ss*, all of which were required to learn the same eight paired-associates by the anticipation method. Three of these groups learned under massed practice (8 sec. between trials) and will be referred to as M-I, M-II, and M-IV. M-I was the standard paired-associate situation, wherein each stimulus term was followed by only one response term. M-II was the same except that each stimulus term was followed by two response terms only one of which was correct. In M-IV, each stimulus term was followed by four response terms, only one of which was correct. The corresponding distributed practice groups (2 min. between trials) will be designated D-I, D-II, and D-IV.

Materials.—The stimulus elements were eight of the nonsense symbols used by Gibson (**8**).[2] The response terms were, depending on the experimental group, 1, 2, or 4 nonsense syllables. These were presented on a paired-associate memory drum in such a manner that the stimulus first appeared alone for 2 sec., then the stimulus and the response (or possible responses) appeared together for 2 sec., then the cycle repeated with a different symbol and different syllables.

All syllables were between 26% and 27% association value.[3] They were selected to be as dissimilar to each other as possible. Of the eight correct syllables, none had the same first letter, nor the same last letter, and none had two letters in common. These were the same syllables for each group. The symbols and their respective syllables appeared in four different sequences to eliminate serial effects. In Groups M-II and D-II, where each symbol was accompanied by two syllables, each syllable appeared in the first position half of the time, each appeared in the second position half of the time. In Groups M-IV and D-IV each of the four syllables accompanying any symbol appeared in each of the four possible positions one-fourth of the time. Thus, if the four possible *R* terms for S_1 first appeared in the order R_{1a}, R_{1b}, R_{1c}, R_{1d}, they might next appear in the order R_{1c}, R_{1a}, R_{1d}, R_{1b}.

Subjects.—Each of the six groups contained 20 *Ss*, selected at random from students in elementary psychology having Ohio State Psychological Examination scores above percentile 30.[4]

[2] The reason for using nonsense symbols for the stimulus terms instead of nonsense syllables was that it was hoped that interference due to the many conflicting nonsense syllables being used in the more complex conditions would be somewhat lessened.

[3] The syllables were taken from lists of nonsense syllables for paired-associate learning studies prepared by A. W. Melton.

[4] In a pilot study, 3 out of 30 *Ss* failed to learn the material. They were the only *Ss* with OSPE's below percentile .30. It was hoped that a matching procedure using the OSPE as a criterion could be used, but after half of the *Ss*

All *Ss* were naive to psychological experimentation, and did not know the purpose of the experiment. Five *Ss* were discarded for failure to learn. All had 80 trials and did not appear to be approaching the criterion. The slowest *Ss* to learn required 56 trials. The five eliminated *Ss* were distributed as follows: M-IV—2, D-IV—2, M-II—1. They were replaced by other *Ss*.

Procedure.—The *Ss* learned the lists to one perfect recitation by the method of anticipation on the first day, returned the next day at the same time, and relearned to the same criterion. Learning was preceded by rather elaborate instructions and demonstration of procedure to compensate for the lack of a practice period. The procedure varied according to distribution in that in the distributed groups, 1 min. and 56 sec. of the 2-min. interval between each trial was filled by naming colors from a large board containing 60 color squares, arranged in ten columns. The *Ss* named the colors in rhythm with a metronome which beat once every 2 sec. until the *Ss'* rhythm was established. Any deviation from the 2-sec. rhythm resulted in the metronome beating again. In the massed condition, the 8-sec. interval between trials was unfilled. The *Ss* were required to spell out the syllables, and the noncorrection method was used to hold the work per trial constant as the situations increased in the degree of discovery required. Evocation of the correct response caused a buzzer to sound, indicating correct, whether the *S* anticipated the correct response or merely said it after it appeared. This was necessary in order to inform *Ss* in the more complex conditions when they had located the correct response. After *S* had anticipated all syllables correctly in the same trial, he was requested not to think about the material until he returned the next day.

Results

As may be seen from Table 1, the D groups took fewer trials to reach the criterion in all three conditions. The over-all F (4.10) between the massed and the distributed conditions is significant at the 5% level of confidence, and the consistency of the differences between the means would support the assumption that the magnitude of the F ratio is due, in part, to the differences between these means. However, as Bartlett's test for homogeneity of variance yields a χ^2 of 21.37 ($p < .01$) and as the group means and group variances do not covary in any systematic way, no statement concerning the probability of such differences occurring can be made. The size of the differences between Groups M-I and D-I, and between these groups and the other groups, indicates the possibility that a large part of the variance is attributable to these differences. The use of t tests between the groups would, if legitimate, support this interpretation, but the fact that the interaction variance does not approach significance makes this interpretation tenuous at best.

TABLE 1

Mean Number of Trials to One Perfect Recitation

Choices	Massed		Distributed	
	Mean	*SD*	Mean	*SD*
I	26.85	9.45	21.05	7.51
II	37.45	13.72	34.00	8.85
IV	40.70	6.32	39.40	7.99

Figure 1 indicates for each group the mean number of trials required to anticipate correctly one nonsense syllable, to anticipate two in the same trial, to anticipate three in the same trial, and so on until all eight nonsense syllables have been anticipated in the same trial. Examination of the figure indicates that the trends of M-I and D-I are divergent, the trends of M-II and D-II are approximately parallel, and that the trends of M-IV and D-IV at first diverge, then converge. This is taken to indicate that as there is an increase in the degree of discovery required, the effectiveness of the distribution of practice appears relatively earlier in the learning process. This seems to be in conflict with

had been run in the experiment, no significant correlation between OSPE's and learning scores emerged, so the selection on a random basis was continued.

the hypothesis that in the early stages of learning, massing of practice should benefit those situations requiring a greater degree of discovery.

In order to analyze more precisely the data with respect to the location of the beneficial effects of distribution, the data were divided into two parts. The first, which will be called the *discovery* period, is, for each *S*, the mean number of trials required to anticipate first each correct nonsense syllable. The second, which will be called the *fixation* period, is obtained for each *S* by taking the number of the trial following the last failure of anticipation for each of the correct syllables, subtracting from this value the corresponding discovery value, and averaging the eight values. This will be recognized as an averaging of the range of oscillations at the threshold of recall used by Hull *et al.* (13), which is presumably an index of the interference between the items in a task (at least in serial learning situations).

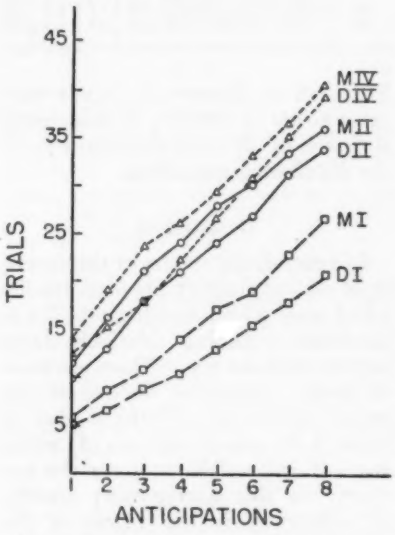

FIG. 1. Mean number of trials to various numbers of correct anticipations within a trial

TABLE 2

MEAN NUMBER OF TRIALS TO THE FIRST CORRECT ANTICIPATION OF EACH SYLLABLE

Choices	Massed		Distributed	
	Mean	*SD*	Mean	*SD*
I	13.55	4.63	10.72	2.92
II	23.38	9.41	20.95	5.60
IV	26.53	5.04	23.84	5.60

Table 2, which shows the results of the analysis of the discovery period, indicates a uniform statistically significant superiority of the distributed conditions. The F of all conditions was 21.80 ($p < .001$); for distribution F was 5.8 ($p < .05$); and for degree of discovery required F was 51.50 ($p < .001$). A series of steps was taken in an effort to determine whether the superiority of the distributed groups was related to differences in the variability of response. The data in the II and IV conditions were analyzed with respect to the number of response shifts from trial to trial, response shifts being defined as any change in the response to a stimulus from one trial to the next, regardless of the degree of the change. The I conditions were not included because the response shifts were negligible. Table 3 indicates that the hypothesis of no difference between M and D conditions cannot be rejected ($F = .01$), and that no evidence for massed practice as a determiner of variability has been found by this procedure.[5]

As a final step in the analysis of the discovery scores, the number of cor-

[5] The possibility that the variability may have operated in some differential way was considered, and response shifts away from the correct responses were counted, but the over-all F had a $p > .05$ and no consistent trends were in evidence.

TABLE 3

MEAN RESPONSE SHIFTS PER SYLLABLE PER TRIAL AND MEAN CORRECT RESPONSES PER EXPOSURE TO FIRST CORRECT ANTICIPATION FOR EACH SYLLABLE

Choices	Massed				Distributed			
	R Shifts/ Syl./Trial		Correct R/ Exposure		R Shifts/ Syl./Trial		Correct R/ Exposure	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
II	.26	.068	.67	.091	.31	.069	.63	.220
IV	.54	.069	.40	.087	.53	.081	.39	.061

rect responses made during the discovery period was divided by the number of trials in the discovery period. These data, also summarized in Table 3, revealed no difference between the means of the massed and distributed conditions, analysis of variance yielding an F of 2.24.

The data from the fixation period are shown in Table 4. The analysis of variance on all groups gave an F of 2.28, which is short of the 5% level of confidence, but is of considerable interest as it indicates the possible locus of the superiority of massed practice over distributed practice. Inspection of the means shows that in the M conditions increasing complexity required fewer fixation trials, while in the D conditions increasing complexity required more fixation trials.

TABLE 4

MEAN NUMBER OF TRIALS FROM THE FIRST CORRECT ANTICIPATION OF EACH SYLLABLE TO THE TRIAL FOLLOWING THE LAST FAILURE TO ANTICIPATE EACH SYLLABLE

Choices	Massed		Distributed	
	Mean	*SD*	Mean	*SD*
I	6.84	3.75	4.05	3.59
II	6.61	4.86	5.72	3.76
IV	5.84	2.82	6.17	3.74

On the basis of these findings, we may conclude that the general superiority of distributed practice is in large part due to effects operating early in the learning process. As practice continues, these beneficial effects of distribution tend to disappear, and it is possible that in the more complex conditions there is actually a reversal, the massed conditions becoming more beneficial.

Retention after 24 hrs. was measured both by recall and by relearning. The results are shown in Table 5.

TABLE 5

MEAN ANTICIPATIONS ON FIRST RETENTION TRIAL AND MEAN TRIALS TO RELEARN TO ONE PERFECT RECITATION

Choices	Correct Responses on First Relearning Trial				Mean Trials to Relearn			
	Massed		Distributed		Massed		Distributed	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
I	4.05	1.53	5.10	1.22	4.6	3.0	2.6	2.7
II	3.35	1.71	4.15	1.94	5.4	3.4	4.2	2.6
IV	3.45	1.99	4.60	1.83	6.0	3.7	4.4	3.6

They fail to disclose significant variation either in recall or in relearning, although in all cases the means favor the distributed conditions.

DISCUSSION

In general, the results of the experiment do not support previous studies which have found massing superior in situations requiring relatively large degrees of discovery. There are three obviously important differences between this study and the studies of Cook (1, 2) and of Ericksen (6) which make it impossible to state the *one* reason for this discrepancy; namely, (a) differences in the degree of distribution, (b) some differences in the degree of discovery (very complex

mazes or puzzle situations presumably requiring a greater degree of discovery than the complex conditions of this study), and (c) differences in the nature of the task. Without considerable investigation, it would be impossible to state the optimal degree of distribution for any given task, but it is possible that some of the experiments referred to found massing superior for the simple reason that the distributed practice period was too distributed, with the negative effects of forgetting the required responses being greater than the positive effects of distribution. Also, extrapolation of the trends of the data in Table 1 and Table 4 indicates the possibility that more complex conditions would result in the superiority of massed practice.

The factor or factors influencing the superiority of the D conditions occurred mainly in the early trials, a finding not in agreement with the assumption of the Ericksen study that early massing would aid in tasks which require relatively large amounts of variable behavior. Furthermore, there was no indication of greater variability in the M conditions than in the D conditions. Although increased variability as a result of massing does occur in some rigidly circumscribed conditions (4, 22), the phenomenon has yet to be demonstrated in situations requiring complex patterns of response which change from trial to trial (21). The additional finding that the M and D conditions do not differ significantly in the number of correct responses per trial in the discovery period seems to indicate that the superiority of the D groups is due to some factor other than an early identification and practicing of the correct response. These findings indicate that the prediction of increased variability and beneficial effects of distribution as a function of the greater accumulation of inhibitory effects with greater massing, as proposed by Hull (12) and as elaborated by Solomon (19), does not extend to the present situation. Although it is recognized that accumulated inhibitory effects might account for the slower speed of anticipation in the M conditions, the fact that the *S*s in the D conditions were continuously responding during rest intervals makes this interpretation dubious in the present case.

Another hypothesis which has been suggested to account for certain beneficial effects of distribution of practice is the interference hypothesis advanced by McGeoch (16), which assumes that weak responses are forgotten more rapidly than strong responses as a result of the interference of other competing responses. As the rest period in distribution creates an opportunity for additional responses to compete with and thus eliminate the weak responses, there should be more correct responses per trial in the D conditions. That this was not true in the discovery period indicates that the differences in performances lay in the stage of learning at which anticipation began, and this difference is not due to a difference in the speed with which the correct response is learned. To state this in a more general way, the superiority of the D groups is revealed in recall but not in recognition.

Although the precise factors associated with this result are not indicated in the present study, Gibson's (7) theory of verbal learning would suggest the possibility that the beneficial effects of distribution would arise either in learning to discriminate the stimuli or in the process of associating each S-term with its R-term. Either process would provide an opportunity for differential forgetting to occur.

One aspect of the data from the discovery scores which requires consideration is the lack of apparent change in the superiority of distributed practice with increases in the degree of discovery required. According to the original hypothesis, there should have been an increase in the effectiveness of massing. According to the interfering-response hypothesis, if increased discovery requirements bring increased interference, then distribution of practice should show an increasing superiority. Neither of these assumptions is supported. Apparently, increasing the number of choices from one to two to four increases the difficulty of discovering the correct response, but does not appreciably affect the amount of intra-list interference. This is also indicated by the trends of the fixation scores. McClelland (15) has reported similar results and has offered an explanation to the effect that generalization between response terms decreases rather than increases when the choice is between right and wrong responses rather than competing correct responses.

Consider the data from the fixation period. As was indicated, the groups do not differ significantly, but the trends are of some interest, and they are consistent with the above explanation of the beneficial effects of distribution. In the discovery period, the superiority of the distributed condition was present in all conditions, but this superiority continued in the fixation period only in the case of M-I and D-I, these groups differing from the other groups in that *S*s in the I groups had less experience with the stim-

ulus terms before they started to anticipate the responses. This occurred because there was less of a problem in learning the correct responses, and anticipations could be made before the stimulus terms were highly differentiated from each other. Thus, some of the task in which distribution was beneficial remained to be done in the fixation period. In the II and IV conditions, the location of the correct responses took progressively more trials; consequently more practice with the stimulus terms occurred in the discovery period and there was less opportunity for beneficial effects of distribution to occur in the fixation period. Another factor which was probably operating to lower the effectiveness of distribution of practice was fatigue. It should be emphasized that the distribution conditions required 2.53 min. from the start of one trial to the start of the next, whereas the massed conditions required .67 min. from the start of one trial to the start of the next. The combination of these two factors seems sufficient to account for the trends during the fixation period. This interpretation would not account for the results of other investigators who found massing superior in complex situations, as the rest periods in these other studies were not filled, and there is no reason to assume that fatigue should have been operating in these cases.

The slight, statistically insignificant, but very consistent mean superiority of the distributed conditions in both recall and relearning does not indicate any change in the relationship between distribution of practice and degree of discovery after 24 hrs. This study does not throw any particular light on the problem of the superiority of distribution of practice in retention, there merely being an indication that this phenomenon apparently exists in more complex conditions as well as in the simplest rote learning situations.

Summary

1. The purpose of the experiment was to investigate the relationship between the number of possible responses to each stimulus and distribution of practice, where the conditions of distribution were comparable to a number of other recent studies in the field of rote learning. Data were collected to test the assumption that massed practice becomes more effective as the number of response choices to each stimulus increases, this increased effectiveness being a function of the increased variability resulting from the massing with a consequent increased efficiency in locating the correct responses.

2. Three groups of *Ss* learned, with massed practice, to give nonsense syllable responses to nonsense symbols, 8 sec. being interpolated between each trial. These groups differed in that only one syllable for each symbol was exposed for Group M-I whereas two syllables appeared with each symbol for Group M-II and four syllables appeared with each symbol for Group M-IV. In each case only one syllable was correct for each symbol, and the task of S was to learn to give the correct syllables by the anticipation method. Three corresponding groups learned with distributed practice (2-min. rest between trials, filled by reading colors from a board). Twenty-four hours after learning to a criterion of one perfect trial, retention was tested by relearning to the same criterion.

3. No evidence for greater variability of responses under either condition was found, nor was there any evidence that the correct responses were more rapidly located or learned more effectively under either condition.

4. The distributed conditions were superior to the massed conditions in the anticipation of the correct responses. This is most strikingly evidenced in the mean number of trials to the first correct anticipation for each nonsense syllable, but is also true in trials to the criterion of mastery. Retention after 24 hrs. shows trends indicating the same consistent superiority of the distributed conditions.

5. The major beneficial effect of the distribution of practice comes early in the learning process. There are trends which are not statistically significant that indicate a possibility

that massing may be beneficial late in learning when there are a number of alternative responses.

(Manuscript received April 16, 1951)

References

1. Cook, T. W. Massed and distributed practice in puzzle solving. *Psychol. Rev.*, 1934, **41**, 330–355.
2. Cook, T. W. Factors in massed and distributed practice. *J. exp. Psychol.*, 1944, 34, 325–334.
3. Dennis, W. A comparison of the rat's first and second explorations of a maze unit. *Amer. J. Psychol.*, 1935, **47**, 488–490.
4. Dodge, R. *Conditions and consequences of human variability.* New Haven: Yale Univ. Press, 1931.
5. Ericksen, S. C. The effect of distributed and massed practice on the nature of errors made by normal and operated rats. *J. comp. physiol. Psychol.*, 1941, **31**, 57–68.
6. Ericksen, S. C. Variability of attack in massed and distributed practice. *J. exp. Psychol.*, 1942, **31**, 339–345.
7. Gibson, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.
8. Gibson, E. J. Retroactive inhibition as a function of the degree of generalization between tasks. *J. exp. Psychol.*, 1941, **28**, 93–115.
9. Heathers, G. L. The avoidance of repetition of a maze reaction in the rat as a function of the time intervals between trials. *J. Psychol.*, 1940, **10**, 359–380.
10. Hovland, C. I. Experimental studies in rote learning theory. V. Comparison of distribution of practice in serial and paired-associate learning. *J. exp. Psychol.*, 1939, **25**, 622–633.
11. Hovland, C. I. Experimental studies in rote learning theory. VIII. Distributed practice of paired associates with varying rates of presentation. *J. exp. Psychol.*, 1949, **39**, 714–718.
12. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
13. Hull, C. L., Hovland, C. I., Ross, R. T., Hall, M., Perkins, D. T., & Fitch, F. B. *Mathematico-deductive theory of rote learning.* New Haven: Yale Univ. Press, 1940.
14. Lashley, K. S. A causal factor in the relation of the distribution of practice to the rate of learning. *J. Anim. Behav.*, 1917, **7**, 139–142.
15. McClelland, D. C. Studies in serial verbal discrimination learning. I. Reminiscence with two speeds of pair presentation. *J. exp. Psychol.*, 1942, **31**, 44–56.
16. McGeoch, J. A. *The psychology of human learning.* New York: Longmans, Green, 1942.
17. McGinnis, E. The acquisition and interference of motor habits in young children. *Genet. Psychol. Monogr.*, 1929, **6**, 203–309.
18. Melton, A. W. Learning. In W. S. Monroe (Ed.), *Encyclopedia of educational research.* (2nd Ed.) New York: Macmillan, 1949.
19. Solomon, R. L. The influence of work upon behavior. *Psychol. Bull.*, 1948, **45**, 1–40.
20. Solomon, R. L. A note on the alternation of guesses. *J. exp. Psychol.*, 1949, **39**, 322–326.
21. Thorndike, E. L. The refractory period in associative processes. *Psychol. Rev.*, 1927, **34**, 234–236.
22. Wingfield, R. C., & Dennis, W. The dependence of the rat's choice of pathways upon the length of the daily trial series. *J. comp. Psychol.*, 1934, **18**, 135–147.

THE DECLINE OF PITCH DISCRIMINATION WITH TIME

J. DONALD HARRIS

U. S. Naval Medical Research Laboratory

Interest in the matter of pitch discrimination has been increasing sharply in very recent years, with developments in our knowledge of how the auditory nerve responds (and even how its individual fibres respond) to pure tone stimulation, and with developments in various techniques for studying the sharpness of tuning of the basilar membrane.

In this paper are reported experiments in which *O* hears two tones separated in time, and is required to judge the second "higher" or "lower." The question is whether the judgment on the second tone is always referred to some relatively stable standard resulting from the first tone, even one presented many hours previously, or whether pitch discrimination demands close contiguity of the two stimuli in time so that interpretation can properly utilize such neural concepts as delay circuits, inhibition, and summation. If discrimination does not decline during a period of time longer than the immediate neural events brought on by the first tone, then those immediate neural events are inappropriate as an explanation.

The effect of time on judgments of pitch has been the subject of several papers. The earlier ones were interested largely in a study of the "memory-image," while the later ones were interested more in the changes in the time error than in the changes in sensitivity.

The most illuminating study for the present purpose is, oddly enough, the first one, in 1886, by Wolfe (9). He used impure tones of 1-sec. duration as standard stimuli, and comparison stimuli differing by 0, 4, 8, and 12 cps. The inter-stimulus intervals ranged up to 180 sec., the larger part of the data being collected at 60 sec. and less. It is important to note that Wolfe guarded against forming a sense of absolute pitch, by using up to 11 different frequencies interchangeably as standards. Later workers generally have disregarded this vital factor. Wolfe found no clear difference among the intervals from 1 to 5 sec., but subsequently a very slow decline in sensitivity occurred up to 120 sec., the limit of the experiment. Wolfe fitted the curves for three *O*s by the equation:

$$\text{Per cent correct} = \frac{kf}{\log t} + cf,$$

where f = frequency, t = time, and k and c are constants.

Using the same instrument to produce clangs, Angell and Harwood (2) duplicated in 1899 much of Wolfe's experiment. Their results confirmed Wolfe's study in general although there were distinct differences in initial ability, and in resistance to decline of sensitivity, between the two groups of *O*s. A repetition of Angell and Harwood, with the same instrument, was done by Whipple (8), with essentially the same results.

In 1914 Anderson (1) studied inter-stimulus intervals of $\frac{1}{16}$, $\frac{1}{8}$, $\frac{1}{4}$, $\frac{1}{2}$, 1, 2, 3, and 4 sec. with a view to determining the most efficient one for discrimination. He concluded that "the intervals tested are practically equally favorable, with possibly a slight tendency in favor of the shorter."

Irwin (5) in 1937 used intervals of 1, 10, 30, and 60 sec., with carefully controlled 256-cps tuning forks with acoustic filters. He reported changes of differential threshold with respect to the 1-sec. interval, and found an average decline from 1 to 10 sec. of .43 cps. At 30 and 60 sec., the data revealed no further regular decline.

Koester (6) directly attacked the problem in 1945 with quite adequate apparatus and technique, but covered such a range of stimuli that on any particular point his data are rather scanty. His results with four practiced *O*s with use of the constant method showed no difference in threshold between the zero and 1-sec. intervals, and an average increase in the size of the variable error of about .8 cps from 1 to 10 sec. His results with 14 unpracticed *O*s at the intervals 1, 3, 6, and 9 sec. were quite inconclusive. Even at the shortest interval and at the easiest level, 12 cps, this group was not

performing better than chance so that it was not possible to determine deterioration as a function of time.

For intervals longer than 10 sec., Koester chose for each of three practiced *O*s a difference detected only about 80% of the time at zero interval, and presented the two stimuli in a random sequence, giving a tone every 5, 15, or 47 sec. in three different series. No significant difference occurred up to 15 sec., but deterioration was evident at 47 sec.

Postman (7) is the latest to investigate the inter-stimulus interval for pitch, seeking to measure the time error, if any. He obtained average DL's for nine *O*s at 0, 1, 2, 4, and 6 sec., at each of seven frequencies. The variability shown in his results makes it difficult to see any influence of interval. At only three frequencies was the 0-sec. interval the most efficient, and at only two frequencies was the 6-sec. interval the least efficient.

Apparatus and Procedure

The outputs of two sine-wave oscillators were led to the two channels of a "clickless" electronic switch, and then to two attenuators, a mixing circuit, and a group of 12 matched phones. One oscillator (General Radio Model 913C) had been modified so that it could be varied in half-cycle steps. The timing control of the switch passed one tone for 1.4 sec., allowed an interval of silence from .1 to 25 sec., then passed a second tone of 1.4 sec. The timing was performed by a Potter Instrument Company electronic device, accurate to a millisecond. Small frequency steps on the variable oscillator were calibrated with a Conn chromatic stroboscope. Zero differences in frequency between the oscillators were obtained by use of the null point meter provided on one of the oscillators (General Radio Model 617C).

The procedure was a modification of the method of constant stimulus differences, using two categories, "higher" and "lower." The standard stimulus, at 1000 cps, was always given first. Comparison stimuli were chosen so that at least ten points were determined within the useful range of the psychometric function. The constants of the psychometric function for each *O* were determined using the Müller-Urban weights. All functions were based on at least 360 judgments, i.e., 36 judgments per comparison stimulus.

A 40-item practice series was given to each *O*. Those who scored 11 or more items incorrectly (about 5% of the initial group of *O*s) were not used further, since they were so poor in discrimination that no threshold could have been computed for them when using the particular comparison stimuli chosen. In the experiment proper, systematic practice effects were minimized by randomizing the inter-stimulus intervals. Since *O*s were not available for an unlimited time, it was not possible to use each *O* to explore all the inter-stimulus intervals desirable. It was necessary to rely upon large numbers. A total of 213 young men with normal hearing acuity was used, in groups of 12 or fewer.

The effect on pitch discrimination of elapsed time between stimuli has been shown in a previous report (3) to depend to a considerable extent upon whether the standard stimulus is always of the same frequency. When it is, *O* can quickly build up from the preceding stimuli a subjective standard upon which he bases his judgment of succeeding stimuli. This subjective standard remains stable for many seconds. Consequently, discrimination with a single standard stimulus is relatively little affected by the time between standard and comparison stimuli.

If, on the other hand, the standard stimulus is changed in frequency for every judgment, then the range of the standard and variable stimuli is considerably greater, the preceding sequence of stimuli furnishes only a relatively coarse anchor for judging, and *O* is forced to consider only the two particular stimuli of a comparison pair. Under these conditions, it is possible to show that time has a measurable effect on discrimination.

Accordingly, in this experiment two procedures were used, one with a fixed standard and one with a standard varying from 950 to 1050 cps, in 5-cps steps. It has been shown by using the method of single stimuli that the latter procedure provides a subjective standard so coarse that the DL is greater than 30 cps (4).

Results and Discussion

Because of the greater practical importance of pitch discrimination at inter-stimulus intervals of a very few seconds and less, these times were sampled especially carefully. Table 1 contains data from .3 to 3.5 sec. No difference was found within this period when a fixed standard was employed. This corroborates the data of Anderson (1), who found no reliable difference on the average or for any of four *O*s at the intervals $\frac{1}{16}$, $\frac{1}{8}$, $\frac{1}{4}$, $\frac{1}{2}$, and 1 sec.

For the longer intervals it was

TABLE 1

MEAN DIFFERENTIAL PITCH THRESHOLDS IN CPS WITH VARIOUS BRIEF INTERVALS BETWEEN TWO TONES

Group	N	Inter-Stimulus Interval in Seconds					
		.3	.5	1	2	3	3.5
A	35			6.1	6.3	5.9	
B	68	5.5			4.6		
C	75	5.1		4.75			
D	59	5.6					5.0
E	88		4.75		4.7		
F	87		4.65	4.9			
G	35				4.9		4.7

possible to use the same *O*s with all of the intervals. One group of 32 *O*s was tested with intervals of 3, 7, 15, and 25 sec. with a fixed standard, another group of twelve *O*s was tested with intervals of .1, 1, 3, 7, and 15 sec. with the roving standard. The data from these experiments are summarized in Table 2 and Fig. 1. The data from the tests with a fixed standard corroborate in general Koester's (6) results at 0, 1, 5, and 10 sec. He found an average increase of .8 cps after 10 sec.; an interpolation at 10 sec. in our data gives an increase in the DL of .65 cps. Differences in the methods employed or individual differences in *O*s could well account for the disparity. Unfortunately, the earlier writers who used a roving standard did not report their results in terms of a differential threshold, so that further exact comparisons with previous work cannot be made.

When the two curves of Fig. 1 are compared, a considerable difference in effect of elapsed time is clear. With the fixed standard, we appear to be studying the relatively slow deterioration of the subjective standard, which is so stable that discrimination has declined by less than 1 cps after 15 sec. On the other hand, with the roving standard we are studying the relatively rapid deterioration of the after-effects of the standard stimulus alone. This process results in the same decline of discrimination after only 4 sec.

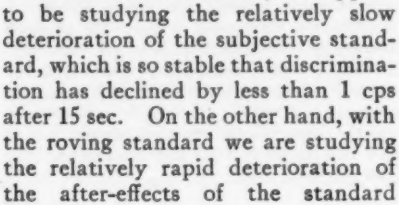

FIG. 1. Pitch sensitivity at 1000 cps with relatively longer inter-stimulus intervals

TABLE 2

MEAN DIFFERENTIAL PITCH THRESHOLDS IN CPS WITH VARIOUS RELATIVELY LONG INTERVALS BETWEEN TWO TONES

Inter-Stimulus Interval (in Sec.)	Type of Standard Tone			
	Fixed (N = 32)		Roving (N = 12)	
	Mean	*SD*	Mean	*SD*
0.1			4.2	1.51
1			4.3	1.20
3	3.9	1.15	5.1	1.36
7	4.1	1.06	5.4	1.22
15	4.7	1.43	7.9	1.60
25	6.9	1.81		

Of especial interest is the comparison between intervals .1 and 1 sec. with the roving standard. The longer interval results in a nonsignificant deterioration of only .1 cps. Evidently pitch discrimination may be based upon neural events which are of relatively long duration.

SUMMARY AND CONCLUSIONS

Previous experiments on the effect of elapsed time between the standard and comparison stimuli on pitch discrimination have suffered from either of two defects: (a) insufficient information was obtained to permit

the calculation of differential thresholds; or (b) a simple fixed standard stimulus was used so that a stable subjective anchor could be built up, which permitted *O* to maintain discrimination over a long inter-stimulus interval.

The present experiment showed that with a fixed standard stimulus, no decline in discrimination occurs up to 3.5 sec., a decline of .8 cps occurs after 15 sec., and a decline of 3.0 cps occurs after 25 sec. With a roving standard stimulus, on the other hand, no decline occurs from .1 to 1.0 sec., but delays of 3, 7, and 15 sec. result in changes of .85, 1.1, and 3.7 cps respectively. It is clear that there occurs an appreciably more rapid deterioration of the after-effects of a single stimulus than of the subjective anchor built up from repeated stimuli close together in frequency.

Even with a roving standard, which allows for the formation of only a very coarse subjective anchor, and forces *O* to refer his judgment more completely to the immediately preceding standard stimulus, no real loss of pitch discrimination occurs between the inter-stimulus intervals of .1 and 1 sec. It may be argued that this fact suggests an explanation of pitch discrimination on the basis of some other neural event than a delay circuit, inhibition, or summation. It seems clear that each single stimulus contributes to a subjective standard at a level over and above the neural after-effects on the basis of which a single comparison is made.

(Manuscript received April 13, 1951)

References

1. Anderson, D. A. The duration of tones, the time interval, the direction of sound, darkness and quiet, and the order of stimuli in pitch discrimination. *Psychol. Monogr.*, 1914, **16**, Whole No. 69, 150–156.
2. Angell, F., & Harwood, H. Experiments on discrimination of clangs for different intervals of time. *Amer. J. Psychol.*, 1899, **11**, 67–79.
3. Harris, J. D. Pitch discrimination and absolute pitch. USN Bureau of Medicine and Surgery Research Report, Project NM 003 026, 30 January, 1948.
4. Harris, J. D. Discrimination of pitch: suggestions toward method and procedure. *Amer. J. Psychol.*, 1948, **61**, 309–322.
5. Irwin, C. C. A study of differential pitch sensitivity relative to auditory theory. *J. exp. Psychol.*, 1937, **21**, 642–652.
6. Koester, T. The time error and sensitivity in pitch and loudness discrimination as a function of time interval and stimulus level. *Arch. Psychol., N. Y.*, 1945, **41**, No. 297.
7. Postman, L. The time-error in auditory perception. *Amer. J. Psychol.*, 1946, **59**, 193–219.
8. Whipple, G. M. An analytic study of the memory image and the process of judgment in the discrimination of clangs and tones. *Amer. J. Psychol.*, 1901, **12**, 409–457.
9. Wolfe, H. K. Untersuchungen über das Tongedächtniss. *Philos. Stud.* (Wundt), 1886, **3**, 534–571.

EFFECTS OF FAILURE STRESS UPON SKILLED PERFORMANCE

RICHARD S. LAZARUS AND CHARLES W. ERIKSEN

The Johns Hopkins University

The way in which people perform under conditions of psychological stress is of great interest to psychologists. However, there has been relatively little experimental work on this topic. Not only does this general problem have important applied implications for military psychology, industry, and education, but it is of considerable theoretical importance. Stress situations are useful for the study of various aspects of human learning, personality, and the emotional processes.

The majority of interest in the field seems to have centered around emotional reactions to stress and the physiological consequences of stress conditions. Much of this work has been observational in nature. While the physiological and behavioral manifestations of stress reactions are important, it is at least equally important to study what happens to a person's performance under stress. There has been a tendency to assume that, under stress, performance will either deteriorate or remain unchanged. In fact, most of the experiments that have been done along these lines have produced situations which, if performance was studied, resulted in losses in effectiveness (**2, 11, 12, 15, 16, 17**). We believe that, if conditions allow it, performance should improve for some *S*s and be unaffected or impaired in others.

The study of the effects of stress upon skilled performance offers some rather difficult methodological problems. One of the biggest stumbling blocks to effective research is the difficulty of producing experimentally a realistic threat. One possible criterion of having stressed *S*s is an operational one, that is, producing some change in performance which can only be attributed to the stress condition itself. Many *E*s have simply used situations which were assumed to be stressful. The most frequent techniques have been: (a) making it impossible for *S*s to solve problems which they are motivated to solve (**1, 9, 12, 13, 14**); (b) implying failure to finish in the allotted time (**6**); (c) giving *S*s false failure scores (**4, 6**); and (d) a miscellaneous group of studies involving distractions (**11**), deprivation of desired objects (**3**), threat of electric shock (**16**), disparagement by *E* (**10**), and excess supervision and control of children's play (**5**).[1]

Large gaps exist in our knowledge about stress effects. We know very little about the importance of the type of task. We know little about the importance of when in the learning and performance of a task the stress is imposed and under what conditions of stress *S* must function. Nor do we know much about what components of performance are affected and in what way. We recognize that there are great individual differences in response to stress, but we do not understand the nature of these differences or the probable interactions between the characteristics of *S*s and the conditions of stress.

In the present experiment we have

[1] The authors are endebted to Dr. Jack R. Strange for his assistance in collecting the data.

attempted to demonstrate that a particular situation is stressful in terms of our operational criterion and to analyze the effects upon the speed and accuracy of performance. A later article will present correlations of performance under stress with various personality measures such as the Rorschach test.

Procedure

Our purpose was to produce a situation which would represent a real threat to a group of college students and which would allow opportunity to measure the effects of this threat upon the speed, accuracy, and total efficiency of performance. Two groups were tested. One consisted of male sophomore engineering majors at The Johns Hopkins University. The other was a beginning mixed class in psychology at Western Maryland College.[2] In the analysis of the data, both groups were combined since their results were similar. The total number of cases included 115 experimental *S*s and 73 controls. They were tested during regular class hours.

The *S*s were told that they were participating in an evaluation program, the results of which would be placed on their records. It was emphasized that the intelligence test which was to be administered was an excellent predictor of college success. The *S*s were informed that it was most important for them to do well since their performance might later serve as information for their faculty concerning their status in school.

Both the Western Maryland group and The Johns Hopkins group were administered the following procedure:

After the introductory motivating remarks, an extended version of Form 1 of the Wechsler-Bellevue digit-symbol subtest was given for a 3-min. time period. During the succeeding 1.5 hr. the Group Rorschach test was administered. During this period two *E*s removed the completed digit-symbol tests, presumably for correction.

When the group had finished the Rorschach test, *E* read off a list of names with the instructions that they report to another room. This represented the control group. The experimental and control groups were actually so divided that they were matched with respect to their college grade point averages.

The experimental group was then told in the most severe manner that they had done very poorly on the original digit-symbol test. The seriousness of this failure was reaffirmed and it was suggested that many of them had apparently failed to appreciate the importance to them of this testing. They were told that another administration would give greater certainty that their true ability had been measured and that they would be given one more chance. It was explained that the other group had done satisfactorily on the test.

The experimental *S*s were then given Form 2 of the digit-symbol test in a lengthened version, with a 3-min. time limit. At four intervals during the testing, *E* called out false norms which were beyond the capacities of the students to achieve. Each line of the test had been lettered so that *E* simply said loudly, "You should have finished line B," then later, "You should have finished line E," and so forth. Any protest or comment was met with a sharp, "Quiet."

The *S*s in the control group were told that they had performed well on the digit-symbol test but that it was important to be sure that their true ability had been measured. They were informed that one more measurement would be made. They were told that their final score would be the average of the two testings, and were urged to improve their performance if possible. The attitude of *E* was friendly and encouraging, yet reasonably business-like. Three minutes were allowed for the test, and as in the case of the experimental group, four interruptions were made during the test with the presentations of false norms. However, these norms were well within the level of attained performance of each *S* so that this group was supported rather than threatened throughout the test. Moreover, because of the interruptions the control *S*s had the same amount of distraction as the experimental *S*s.

At the end of the experiment, both groups were told the real nature of the experiment and allowed to give vent to feelings of relief. The purpose of the experiment and its implications were discussed with them.

Results

Table 1 summarizes the total scores of both the experimental and control groups on Test 1 and Test 2. These data make it clear that the essential effect of the stress condition on total performance was to greatly increase

[2] The authors are grateful to Professor Paul Hessemer of The Johns Hopkins University and to Dr. Olive R. Russell of Western Maryland College for their kind cooperation in this experiment.

TABLE 1

MEANS AND VARIANCES OF GROSS SCORES ON TEST 1 AND TEST 2

	Experimental		Control	
	Mean	*SD*	Mean	*SD*
Test 1	111.4	416.2	113.1	436.8
Test 2	109.4	1096.1	109.4	219.0
t	.70		2.5	
F		2.6		2.0
p	.48	<.01	<.02	<.01

the variance. The experimental group variance on Test 2 (the severe stress condition) was significantly higher than on Test 1. The control group, on the other hand, significantly decreased in variance on Test 2 when, for these *S*s, the stressfulness of the situation was reduced. Both groups decreased slightly in mean score. The experimental group's relative loss was less than that of the control group, but was too small to be considered reliable.

In Table 2 are presented the differences in number of errors and speed between Tests 1 and 2. The experimental group shows an increase in errors while the control group shows a decrease, and this difference in the change in errors is statistically significant.

TABLE 2

DIFFERENCES IN ERROR AND SPEED SCORES ON TEST 1 AND TEST 2

Group	Error Scores		Speed Scores	
	Mean	*SD*	Mean	*SD*
Experimental ($N = 115$)	+.643	4.8	−1.29	204.5
Control ($N = 73$)	−.596	4.0	−3.56	112.4
t	3.9		1.3	
F		1.2		1.73
p	<.001	>.05	.20	<.02

Both groups show a slight decrement in speed on Test 2, and the control group's loss is slightly greater than the loss for the experimental group. This picture tends to obscure the true state of affairs, however. Those *S*s who made many more errors on the second testing had a tendency to increase in speed. This compensated for the greater number of errors, and in fact, the increase in speed occurred at the expense of accuracy. Speed accounts for 92% of the variance in total scores. This means that the faster *S* worked, the higher score he made.

As a consequence, the results on speed parallel those on total scores. While both groups show much the same picture with respect to mean change in speed, the variability of the two groups is significantly different.

The effects of stress upon performance are also shown by the correlations between scores on Tests 1 and 2 for the experimental and control groups. These are shown in Table 3. For total scores and speed the correlations for the control group are significantly higher than for the experimental group. Scores on Test 2 are more accurately predicted from Test 1 when stress conditions are not introduced.

All these results indicate that the effect of stress upon performance

TABLE 3

CORRELATIONS BETWEEN FIRST AND SECOND TEST FOR TOTAL SCORES, ERROR SCORES, AND SPEED SCORES

Scores	Experimental Group ($N = 115$)	Control Group ($N = 73$)	t*	p
Total	.645	.901	4.65	.0003
Errors	.183	−.094	1.78	.085
Speed	.687	.870	3.12	.004

* The significance of the differences between r's is evaluated via Fisher's z transformation.

TABLE 4

DIFFERENCES BETWEEN STUDENTS WITH HIGH AND LOW GRADE POINT AVERAGES ON THEIR PERFORMANCE UNDER STRESS

Measure	Grade Point Average		Significance Test	p
	High (N =24)	Low (N =26)		
Mean Stress Score*	+ 6.5	−2.4	t =2.97	<.01
Variance Estimate	54.8	174.8	F =3.19	<.01

* See text.

varies with the individual. The performance of some *Ss* under stress is little affected whereas other *Ss* show large improvement or decrement. Since it was possible that the type of stress effect *S* showed might be related to his grade point average, stress scores were obtained for each *S* in the experimental group. The stress score was obtained by first subtracting the total score on Test 2 from the score on Test 1. This difference score was highly correlated with Test 1 performance ($r = -.69$ for controls and $-.49$ for experimentals). To correct for this correlation, the deviation of each difference score from the value that would be predicted by the Test 1 score was determined. The data from the control group were used in obtaining the prediction equation. The result is a score which has been corrected for Test 1 performance and for which positive values indicate improved performance under stress and negative values indicate a decrement.

In Table 4 are shown these mean stress scores for *Ss* of the experimental group in the lowest and highest quartile in grade point average. The *Ss* with high academic standing did significantly better under stress than did the low group. Also, the *Ss* in the high group were more homogeneous in their performance under stress as is shown by the smaller variance.

DISCUSSION

It must be recognized that the data which have been obtained are specific to the situation (task, instructions, stress inducement, and *Ss*) until proved otherwise. Moreover, the data do not provide information on the "why" of the effects measured. The performance of some *Ss* may improve because of greatly increased motivation aroused by the increased likelihood of failure. Or, in some cases, the motivation may become so great as to be disruptive through anxieties which are generated or accentuated by the situation and which act as an added distraction. It is also possible that some *Ss* did much worse on Test 2 because they became disgusted by the whole situation, or so threatened that their response was to give up or to "psychologically leave the field."

Our data have suggested that *Ss* with low grade point averages are more likely to show impairment and greater variability of performance under stress. Since merely telling people that they are taking an important test has stressful consequences, a certain number of *Ss* may be doing poor college work because of their inability to maintain the performance of which they are capable under such stressful circumstances. Moreover, as a group, *Ss* with low grade point averages might be expected to be more variable since adequate performance under stress is not the sole determiner of academic success. However, an alternative explanation for these findings is possible. *Ss* who have been doing well in college may not be particularly threatened by the experimental situation. Their past success could have given them confidence in their ability. However, interviews with a number of *Ss* after the experiment lead us to believe that our original explanation is the more likely. Even the best *Ss* reported very strong subjective experiences of stress.

In any case, our results do indicate that, at least in the present situation, stress affected *Ss* differently. It causes some *Ss* to do better and others to do more poorly. Simply comparing the means of a control and experimental group may obscure or allow us to overlook this most important finding.

It is interesting to note that the data in Table 1 show that the variability of the control group markedly decreases on Test 2 whereas that of the experimental group shows a large increase. We might think of Test 1 as moderately stressful for both the control and experimental groups (instructions specified that the results of this testing would be entered into the college record). For the control group Test 2 could be considered reassuring because they had been told that they had done well on Test 1. For the experimental group Test 2 was taken with the impression that they had failed Test 1. In view of the nature of these test conditions, we might make the assumption that we had produced three degrees of stress: mild, moderate, and severe. If this is the case we find that the group variability increases with the degree of stress. This notion needs further experimentation.

The situation and task we used were such as to produce an increase in errors under stress. Frequently subjects do well under stress by increasing their speed at the expense of accuracy. In this particular situation, lowered accuracy mattered very little since it was so poorly correlated with total scores. Undoubtedly, however, there are situations where losses in accuracy would be serious. Lindsley (9) finds this kind of reaction to stress but gives it the questionable interpretation of loss in efficiency. One of the general effects of stress may be to push *Ss* to increase both their speed and errors, even when this is not the most efficient thing to do. Much more work needs to be done on this question.

Merely telling people that they are taking a test which will have an important bearing upon their welfare is apt to be a stressful situation. If this is so it has important consequences for testing. Hutt's (8) clinical findings with intelligence testing are added evidence on this point. This kind of situation may, however, be useful in predicting which people will have difficulty with stress situations and which will profit from a degree of stress.

Summary

The effects of a failure threat upon the performance of college students on an extended version of the Wechsler-Bellevue digit-symbol subtest were studied. The main results were:

1. A significant increase in inter-individual variability in test performance occurred in the stress group.

2. An increase in the number of errors made under stress was compensated for to some extent by an increase in speed. Speed scores accounted for 92% of the variance in total scores.

3. Students with high grade point averages showed a tendency to improve their performance under stress. Students with poor academic standing did more poorly and were much more variable.

(Manuscript received April 20, 1951)

References

1. Adams, C. R. Individual differences in behavior resulting from experimentally induced frustration. *J. Psychol.*, 1940, **10**, 157–176.
2. Alper, T. G. Memory for completed and incompleted tasks as a function of personality: an analysis of group data. *J. abnorm. soc. Psychol.*, 1946, **41**, 403–421.
3. Barker, R. G., Dembo, T., Lewin, K., & Wright, M. E. Experimental studies of frustration in young children. In T. M. Newcomb and E. L. Hartley (Eds.), *Readings in social psychology*. New York: Henry Holt, 1947.
4. Diggory, J. C. Responses to experimentally induced failure. *Amer. J. Psychol.*, 1949, **62**, 48–61.
5. Fredericksen, M. An experimental study of the effects of frustration on the behavior of young children, with especial reference to negativism. *Psychol. Bull.*, 1940, **37**, 488. (Abstract)
6. Gates, G. S., & Rissland, L. G. The effect of encouragement and discouragement upon performance. *J. educ. Psychol.*, 1923, **14**, 21–26.
7. Glixman, A. F. Recall of completed and incompleted activities under varying degrees of stress. *J. exp. Psychol.*, 1949, **39**, 281–295.
8. Hutt, M. L. A clinical study of "consecutive" and "adaptive" testing with the revised Stanford-Binet. *J. consult. Psychol.*, 1947, **11**, 93–103.
9. Lindsley, D. B. A study of performance under speed stress. OSRD, 1945, Publ. Bd. No. 1838, Wash., D. C.: U. S. Dept. of Commerce, 1946.
10. McClelland, D. C., & Apicella, F. S. A functional classification of verbal reac-

tions to an experimentally induced failure. *J. abnorm. soc. Psychol.*, 1945, **40**, 376–390.

11. McKinney, F. Certain emotional factors in learning and efficiency. *J. gen. Psychol.*, 1933, **9**, 101–116.
12. Marquart, D. I. The pattern of punishment and its relation to abnormal fixation in adult human subjects. *J. gen. Psychol.*, 1948, **39**, 107–144.
13. Postman, L., & Bruner, J. S. Perception under stress. *Psychol. Rev.*, 1948, **55**, 314–323.
14. Sherman, M., & Jost, H. Frustration reaction of normal and neurotic persons. *J. Psychol.*, 1942, **13**, 3–19.
15. Thorndike, E. L., & Woodyard, E. The influence of the relative frequency of successes and frustrations on intellectual achievement. *J. educ. Psychol.*, 1934, **25**, 241–250.
16. Williams, M. An experimental study of intellectual control under stress and associated Rorschach factors. *J. consult. Psychol.*, 1947, **11**, 21–29.
17. Zander, A. F. A study of experimental frustration. *Psychol. Monogr.*, 1944, **56** (2), 256.

RESPONSE STRENGTH IN A MODIFIED THORNDIKIAN MULTIPLE-CHOICE SITUATION AS A FUNCTION OF VARYING PROPORTIONS OF REINFORCEMENT

ALBERT E. GOSS AND EDWARD J. RABAIOLI
University of Massachusetts

The importance of the frequency and pattern of reinforcements as parameters which influence the acquisition and extinction of responses has been emphasized increasingly in recent experimental (**3, 4, 5, 6, 7, 15, 17**) and theoretical (**8, 10, 14, 16, 18**) literature. However, experimental studies have been restricted to a consideration of either classical and instrumental conditioned responses or simple choice reactions as functions of varying proportions of reinforcement. Further, with the exception of a study by Brogden (**1**), there has been little interest in the more general problem of response strength as affected by varying degrees of post-acquisition changes in the proportions of reinforcement. Accordingly it was the purpose of the present study to investigate (a) acquisition of the relatively complex responses required in a modified Thorndikian multiple-choice situation as a function of varying proportions of randomly administered reinforcement, and (b) modification in the strength of already-acquired choice responses as a function of changed reinforcement proportions, specifically, the complete elimination of reinforcement and the reduction of the proportions of reinforcement to one-half of the corresponding acquisition rates.

Straightforward extension of stimulus response reinforcement principles (**10, 19**) of the relationship between response strength and varying proportions of reinforcement for simple or single-stimulus-unit learning situations to the Thorndikian situation (**19, 20, 21, 22**) generates the prediction that rate of acquisition as measured by trials or correct responses to criterion will be some direct function of the proportion of primary and/or secondary reinforcements.[1] It is doubtful, however, that S–R reinforcement principles for single units should be extended without modification to situations in which each trial consists of the serial and relatively massed intra-trial presentation of several stimulus units. Specifically, it would appear necessary to supplement such principles by an analysis of possible intra-trial and, to a lesser degree, inter-trial interactions among the several stimulus units in terms of principles of response-produced stimulation, stimulus compounding and patterning, and stimulus generalization (**10, 15, 17, 18**). Therefore, in addition to the proportion of reinforcements, variables such as the number and composition of the stimulus units and the time intervals between units and between trials would seem relevant. Further, since a source or sources of secondary reinforcement may develop on the basis of and in addition to the original secondary reinforcers of human learning tasks (for

[1] Humphreys' (**11, 12**) *expectancy* theory has not been developed in a form that permits fairly rigorous experimental tests. Further, it was difficult to decide whether or not Tolman (**23**) would consider reinforcement of relevance for response strengthening in a modified Thorndikian situation. Accordingly, the derivation of predictions was within the framework of S–R reinforcement principles.

example, *E*'s "right") the possible influence of this factor should also be assessed.

Although necessary considerations for a complete theory of the relationship between learning and the proportion of reinforcement, such analyses appear premature in view of the lack of related data on the level of complex responses. Bearing in mind the necessity for future refinements, the present study is considered to be an exploration of the more general hypothesis stated above, namely, that rate of learning increases as some direct function of the proportion of reinforcement.

The derivation of predictions of the effects of changed reinforcement conditions also rests upon considerations regarding the influence of additional sources of secondary reinforcement and the role in stimulus unit interaction of principles of response-produced stimulation, stimulus compounding, and stimulus generalization. Elimination of reinforcement will result in removal from the stimulus compounds antecedent to each choice reaction of stimuli produced by responses to reinforcers. Also, the stimulus effects of responses to nonreinforcement would be added to the compounds. Therefore, the increasing dissimilarity of acquisition and changed reinforcement stimulus compounds resulting from changes in the components of the compounds should lead, by the principle of generalization (**10**), to a decrement in the number of correct responses. Also, since the degree of dissimilarity is presumably greatest for high acquisition reinforcement proportions and least for low acquisition reinforcement proportions, a direct relationship between size of decrements and the proportion of acquisition reinforcements would be anticipated.

In the case of halved reinforcement conditions the effects of increased dissimilarity of acquisitions and changed reinforcement stimulus compounds should be counterbalanced in part by continued reinforcement, although at reduced rates. Accordingly, while dissimilarity of pre-choice stimulus compounds might occasion initial decrements in the frequency of correct responses, the influence of the latter factor should eventually check and reverse those decrements. Because of continued reinforcement and further because of the relatively smaller changes in compound similarity, somewhat smaller decrements for halved than for elimination of reinforcement conditions would be predicted.

Method

Subjects.—Ninety-two men and women from University of Massachusetts undergraduate psychology classes volunteered to serve as *Ss*. Because the experiment involved color discrimination, data for two color-blind *Ss* were eliminated, thus leaving 90 *Ss*.

Apparatus and stimulus materials.—A memory drum was used to present eight modified multiple-choice units (**19, 20, 21, 22**) each of which consisted of three nonsense syllable choice-stimuli underscored with a narrow strip of Milton-Bradley Bulls-Eye construction paper. The 24 syllables had been selected at random from a subgroup of every other one of Witmer's (**24**) consonant nonsense syllables of 8, 13, 17, 21, and 25% association value. They were assigned at random to any one of the eight units, subject to the restriction that each unit consist of syllables of approximately equal association value and have no letters in common. The three syllables of each unit were typed in capital letters on a single line and underscored with one of eight $1\frac{1}{2} \times \frac{1}{4}$-in paper strips of different and easily discriminable colors.

Since Thorndike and his co-workers had not randomized the positions of choice stimuli, choice positions within each unit were not varied in the present study.[2] In order to minimize the influence of initial position preferences on cor-

[2] The color stimuli were introduced to make the stimulus units more distinctive. For a given *S*, correct responses for a given unit could be based entirely on syllable position.

TABLE 1

SUMMARY OF EXPERIMENTAL PROCEDURE

Group	*N*	Acquisition Reinforcement Proportions	Changed Reinforcement Proportions
100–0	15	100	0
100–50	15	100	50
75–0	15	75	0
75–37.5	15	75	37.5
50–0	15	50	0
50–25	15	50	25

rect syllable selection three sequences of syllable-reinforcement conditions were prepared and assigned to *S*s in such a manner that (a) for each of the three sequences there were choices at each of the three possible syllable positions, and (b) each unit's three syllable-position combinations were correct for one-third of the *S*s of each experimental group. Possible serial learning effects were minimized by the use of two randomly determined orders of the same eight units.

Presentation of the units was at a 3-sec. rate with a 10-sec. inter-trial interval.

Procedure.—The experimental design is outlined in Table 1. The *S*s were randomly assigned to one of the six conditions as they reported to *E*. After being tested for color-blindness by means of the Ishihara cards, they read appropriate instructions along with *E*.

During the acquisition sequence *S*s of each group were required to select and to call out the letters of one of the three syllables of each unit. Correct choices as designated by *E* were reinforced immediately by means of a buzzer. The acquisition trials were terminated when *S*s attained a criterion of 15 of 16 correct choices on two successive trials. Details of the acquisition reinforcement procedures for each group were as follows:

100% reinforcement.—The buzzer sounded each time *S* selected and called out the correct syllable.

75% reinforcement.—The sequence of buzzer occurrences for correct choices was randomly predetermined in such a manner that on a given trial six of eight of the correct choices would be reinforced and for each successive block of four trials three of four of the correct choices for a given stimulus unit would be reinforced.

50% reinforcement.—Subject to the restriction of no more than two successive possible reinforcements or non-reinforcements for correct choices for a given stimulus unit, the sequence of buzzer occurrences was randomly predetermined in such a manner that on a given trial four of eight of the correct choices would be reinforced and over several trials 50% of the correct choices for a given stimulus unit would be reinforced.[3]

Without further instructions to *S*, the 20 trials under changed reinforcement conditions were initiated immediately following the second criterion trial. Reinforcement for correct responses was completely eliminated for Groups 100–0, 75–0, and 50–0. The proportions of randomly predetermined reinforcements for each trial and for the correct syllable of each unit were reduced from 100 to 50%, from 75 to 37.5%, and from 50 to 25% for Groups 100–50, 75–37.5, and 50–25 respectively.

Both correct responses and any incidental remarks made by *S* during the experiment were recorded. Before leaving the experimental situation *S* was given a questionnaire which was designed to elicit supplementary information regarding both bases of syllable choice and reactions to changed reinforcement conditions.

RESULTS

Acquisition.—Means and *SD*'s of trials, correct responses and reinforced correct responses to the criterion for the six conditions are presented in Table 2. Statistical analysis by means of the *t* and *F* tests indicated that the differences between the means and variances of trials, correct responses, and reinforced correct responses to the criterion for each of the three pairs of groups which had had the same acquisition reinforcement proportions were not statistically significant. Accordingly, data for the two 100%, the two 75%, and the two 50% groups were pooled. These data are also summarized in Table 2.

Tests of the hypothesis of no differences among mean trials and mean correct responses to criterion for the combined 100, 75, and 50%

[3] The present procedure, which involves variation in the proportion of reinforcements for a single unit over a series of trials, should be distinguished from the procedure of varying the proportion of trials on which *all* responses were reinforced or non-reinforced. It is planned to use the latter procedure in a subsequent study.

TABLE 2

TRIALS, CORRECT RESPONSES, AND REINFORCED CORRECT RESPONSES TO CRITERION FOR EACH ACQUISITION–CHANGED REINFORCEMENT GROUP AND FOR POOLED GROUPS WHICH HAD THE SAME AQUISITION REINFORCEMENT PROPORTIONS

Group	Trials to Criterion		Correct Responses to Criterion		Reinforced Correct Responses to Criterion	
	M	*SD*	*M*	*SD*	*M*	*SD*
100–0	22.8	5.8	98.5	25.8	98.5	25.8
100–50	21.7	9.5	91.7	33.1	91.7	33.1
75–0	33.1	9.4	145.7	36.5	108.9	27.4
75–37.5	37.8	6.0	162.5	39.8	120.2	29.1
50–0	42.9	7.6	179.6	37.2	87.2	19.3
50–25	42.0	11.4	176.7	47.4	87.1	23.5
100	22.3	7.9	95.1	29.8	95.1	29.8
75	35.4	8.2	154.1	39.0	114.5	28.8
50	42.4	9.7	178.1	42.6	87.1	21.5

reinforcement conditions yielded *F*-ratios of 28.8 and 37.47 respectively.[4] Since both values were significant at beyond the 1% level of confidence for 2 and 87 *df*, it may be concluded that whether measured in terms of trials or correct responses to criterion, the 100% or continuous reinforcement group was superior to the 75% group whose performance in turn exceeded that of the 50% group.

The *F*-value of 7.84 for means of reinforced correct responses to criterion was also significant ($p < .01$; $df = 2,87$). However, mean differences for this measure were less striking and, because the 50% condition required the smallest number of reinforced correct responses, in different directions than condition differences for trials and correct responses.

Since there were within-group and between-group differences in correct responses to the criterion, the Hilgard and Campbell (9) modification of Vincent's procedure was utilized to determine and to compare the learning functions of the three groups. When per cent of total correct responses within successive tenths of trials to criterion was plotted as a function of successive tenths to criterion, visual inspection of the three resultant curves suggested that they could be represented by the same approximately linear function.

Changed reinforcement.—Figure 1, whose ordinate values are based on the single trial maximum of eight correct responses, is a graphic representation of mean correct responses for each of the six conditions for the two acquisition criterion trials and for five successive four-trial blocks under changed reinforcement conditions; the points labeled "totals" represent mean correct responses for the 20 changed reinforcement trials as a whole divided by 20. Inspection of the six curves reveals initial downward trends from criterion values to Block II which are somewhat greater for the elimination of reinforcement conditions; these decrements are followed by upward

[4] Bartlett's χ^2 test (2) was utilized to test the homogeneity of the variances of trials, correct responses, and reinforced correct responses. Since the χ^2 values of 1.44, 3.10, and 3.50 for trials, correct responses, and reinforced correct responses, respectively, did not reach the 5% level ($df = 3$), the hypothesis of no differences among group variances could not be rejected.

FIG. 1. Mean correct responses per trial for criterion trials, five successive four-trial blocks, and for the 20 changed reinforcement trials as a whole

somewhat positively accelerated trends from Block II to Block V.

The one-tailed *t* was utilized to test the null hypothesis with respect to intra-group decrements from criterion mean values to means for totals and to means for Blocks I, II, and V. Since the *t*'s for decrements from criterion to total means, which ranged from 2.42 to 6.01, were significant at from the 3% to beyond the 1% levels, it was concluded that for each of the six changed reinforcement conditions there was a significant reduction from criterion levels in the mean frequency of correct responses for the 20 trials as a whole.[5] With the exception of the criterion—Block I comparison for the 75–37.5 group, statistically significant (*t*'s from 2.12 to 8.36 with corresponding *p*'s ranging from the 4% to beyond the 1% level) decrements from criterion to Blocks I and II were obtained. That these decrements were not permanent, however, is suggested by the less significant differences (*t*'s from 1.44 to 3.00 with corresponding *p*'s from 12% to 1%) between criterion and Block V means. Also, although the increments were not statistically significant, the fact that with the exception of the 75–37.5 group, Block V means for each group were higher than corresponding values for Block I adds support to this suggestion.

Analysis of variance for related measures (2) was utilized to test the hypothesis of no differences among the five four-trial block means for each of the six groups. Since the *F*-ratios for Groups 100–50, 75–0, 50–0, and 50–25 (3.75, 3.17, 10.72, and 4.15, respectively) were significant at from the 5% to beyond the 1% level for 4 and 56 *df*, it was concluded that the observed downward-upward trends for these conditions could not be at-

[5] Because the small variances of the criterion trial scores led to marked departures from the homogeneity of the variances of the distributions which were compared, *p* values for the preceding and subsequent *t*'s were based on 7 *df* (2).

tributed to chance fluctuations. The *F*'s for the two remaining groups were not significant.[6]

The null hypothesis with respect to mean correct responses for the 20 trials as a whole (totals) and to four-trial block trends for the six combinations was also tested by means of an analysis of variance technique (2).[7] Failure to obtain a significant *F* (Table 3) for the ratio of the between-group totals variance to the variance estimate based upon *S*s of the same group suggests that the six groups did not differ significantly in mean number of correct responses for the 20 changed reinforcement trials. The ratio of the variances based upon the interaction of methods and trials and the interaction of pooled *S*s and trials permits a test of the hypothesis of no differences among the four-trial block changed reinforcement trends for the six groups. The *F* of .809 which was computed from this ratio was consistent with the hypothesis of no trend differences, that is, it indicated that the changed reinforcement curves for the six groups had similar forms.

Questionnaire.—The questionnaire results indicated that a similar number of *S*s in all groups tended to use color and/or position and letters of the syllable as the bases of choice reactions. Spontaneous incidental remarks made by a few *S*s during changed reinforcement trials, as well as questionnaire results, indicated that some *S*s in all groups recognized correctly that changes in the rate of buzzer presentation had occurred. Specifically, 42 of 45 *S*s in the three elimination of reinforcement groups and 33 of the *S*s of the halved reinforcement groups were able to recognize the changes correctly. A χ^2 test of independence of degree of change and recognition of change yielded a value of 6.48 which is significant at the 2% level for 1 *df*. It was concluded, therefore, that the halved reinforcement groups did not recognize the changes in rate as readily as the elimination of reinforcement groups.

TABLE 3

ANALYSIS OF VARIANCE OF CORRECT RESPONSES UNDER CHANGED REINFORCEMENT PROPORTIONS

Source of Variation	*df*	Variance	*F*
Between-group totals	5	984.07	1.61
Between-subjects in same group	84	114.55	
Total between subjects	89		
Between four-trial blocks	4	171.96	21.05*
Groups × trials	20	6.61	0.81
Pooled subjects × trials	336	8.17	
Total within-subjects	360		
Total	449		

* Significant at the 1% level of confidence.

DISCUSSION AND CONCLUSIONS

Acquisition.—Proceeding from the stimulus-response reinforcement conception of response strength as a continuous and cumulative function of the number of reinforcements, it was predicted that in a modified Thorndikian situation rate of acquisition as measured by the number of trials and/or correct responses to criterion would be a direct function of the proportion of reinforcements. The acquisition data appear to be consistent

[6] Since the χ^2 values for the 75–37.5 and 50–25 groups permitted the rejection of the hypothesis of homogeneous variances at the 1% level, the *F*'s for these groups should be evaluated conservatively. It should be noted, also, that because the χ^2 values were based upon correlated measures, an underlying assumption was violated.

[7] Bartlett's test yielded χ^2 values of 18.49 for totals ($df = 5$), and 115.73 for the 30 four-trial block values ($df = 29$) which were significant at the 1% level. Also, it should be commented that, contrary to underlying assumptions, the χ^2 values for the trial and block variances were based upon correlated measures.

with this prediction, with, specifically, the 100% groups requiring approximately three-quarters and one-half as many trials and correct responses to criterion as the 75% and 50% groups, respectively. Also consistent with the conception of the continuous and cumulative influence of the number of reinforcements was the finding that the Vincent curves for the three acquisition reinforcement schedules were similar in form. Pertinent to and qualifying the preceding conclusion, however, is a reiteration of the necessity for explicitly incorporating principles of the interaction of stimulus units into a more complete systematization of learning in the Thorndikian situation under varying schedules of reinforcement and conditions of unit presentation.

Changed reinforcement proportions.—The curves obtained for changed reinforcement conditions exhibited initial downward trends from criterion levels to the second four-trial block which were followed by upward somewhat positively accelerated trends; in general, the observation of the downward-upward curve forms was supported by statistical analyses. The apparent positive acceleration of the upward trends when combined with the findings of higher means for Block V than for Block I suggests not only that the initial decrements were not permanent but also that had additional reinforcement trials been given, all groups might have reached correct response levels which did not differ significantly from corresponding criterion values. It is possible, of course, particularly in the case of the elimination of reinforcement groups, that additional changed reinforcement trials would have led to a second downward trend.

Contrary to pre-experimental predictions, however, were the reversals of the initial downward trends of the elimination of reinforcement groups; a finding which raises a problem of considerable theoretical interest, namely the identification of a source(s) of continuing reinforcement for the presumably non-reinforced choices of these groups. In this connection it would appear possible that during the acquisition trials the correct syllables themselves had become secondary reinforcers and therefore, because insufficient non-reinforcement trials had been given to extinguish their secondary reinforcing properties, continued to reinforce correct choices. However, the essentially linear relationship between trials and errors to criterion and the proportion of reinforcement as well as the similar Vincent curves are in opposition to the notion that the correct syllables acquired reinforcing properties during the acquisition trials.

A second possibility is based on the assumption that during changed reinforcement trials the selection of other previously incorrect but also non-reinforced choices resulted in greater frustration than the continuation of the previously correct choices. Accordingly, the hypothesis is advanced that as changed reinforcement trials continued, failure to select previously incorrect choices actually occasioned a relative reduction in frustration which served to reinforce the correct choices of the acquisition trials.

A second problem of theoretical interest stems from the Grant, Riopelle, and Hake (5) suggestion that the trend reversal which they observed could be interpreted as compensation for response suppression during the acquisition trials. While the requisite identifications of a source of suppression and of suppressed responses in the multiple-choice situation has as yet not been

possible, this explanation of the upward trend may prove relevant to the present data.

Summary

Ninety undergraduate *Ss* were required to learn a list of eight modified multiple-choice units to a criterion of 15 correct of 16 choices. The units of the list consisted of a colored paper stimulus and three nonsense-syllable choice stimuli and were presented on a memory drum at a 3-sec. rate. A buzzer served as the "reinforcer" for correct responses. The *Ss* were randomly assigned to one of six numerically equal groups. During the acquisition trials, correct choices for two groups were reinforced 100% of the time, two other groups were given reinforcement for 75% of correct responses, and the two remaining groups received 50% reinforcement. Upon completion of the acquisition trials, reinforcement was halved for one of the 100% acquisition groups and eliminated for the other 100% group. The same procedure was followed for the two 75% acquisition groups and for the two 50% groups. All groups were given 20 trials under the changed reinforcement conditions.

The mean trials to criterion of 22.3, 35.4, and 42.4 for the pooled 100%, 75%, and 50% acquisition conditions, respectively, and the parallel relationship among mean number of errors to criterion, were in accordance with the direct relationships between response strength and reinforcement proportions generally observed for conditioned or simple choice responses and thus supported the S–R reinforcement conception of the continuous and cumulative influence of the number of reinforcements. The downward-upward trends which resulted from halving the proportions of reinforcement were in line with the notion of decrements resulting from dissimilarity of acquisition and changed reinforcement stimulus patterns which were offset by continuing reinforcement. The downward-upward trends of the groups for which reinforcement was eliminated were not predicted. It was hypothesized that these trend reversals might have reflected the strengthening effects of unanticipated but continuing sources of secondary reinforcement.

(Manuscript received April 23, 1951)

References

1. Brogden, W. J. The effect of frequency of reinforcement upon the strength of conditioning. *J. exp. Psychol.*, 1939, **24**, 419–431.
2. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
3. Dinsmoor, J. A. The effect of periodic reinforcement of bar-pressing in the presence of a discriminative stimulus. *J. comp. physiol. Psychol.*, 1951, **44**, 354–361.
4. Grant, D. A., Meyer, D. R., & Hake, H. W. Proportional reinforcement and extinction of the conditioned GSR. *J. gen. Psychol.*, 1950, **42**, 97–101.
5. Grant, D. A., Riopelle, A. J., & Hake, H. W. Resistance to extinction and the pattern of reinforcement: I. Alternation of reinforcement and the conditioned eyelid response. *J. exp. Psychol.*, 1950, **40**, 53–60.
6. Hake, H. W., & Grant, D. A. Resistance to extinction and the pattern of reinforcement: II. Effect of successive alternation of blocks of reinforced and unreinforced trials upon the conditioned eyelid response to light. *J. exp. Psychol.*, 1951, **41**, 216–220.
7. Hake, H. W., Grant, D. A., & Hornseth, J. P. Resistance to extinction and the pattern of reinforcement: III. The effect of trial patterning in verbal "conditioning." *J. exp. Psychol.*, 1951, **41**, 221–225.
8. Hilgard, E. R. *Theories of learning.* New York: Appleton-Century, 1948.
9. Hilgard, E. R., & Campbell, A. A. Vincent curves of conditioning. *J. exp. Psychol.*, 1937, **21**, 310–319.
10. Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

11. Humphreys, L. G. The effect of random alternation of reinforcement on the acquisition and extinction of conditioned eyelid reactions. *J. exp. Psychol.*, 1939, **25**, 141–158.
12. Humphreys, L. G. Acquisition and extinction of verbal expectations in a situation analogous to conditioning. *J. exp. Psychol.*, 1939, **25**, 294–301.
13. Jenkins, W. O., & Stanley, J. C., Jr. Partial reinforcement: a review and critique. *Psychol. Bull.*, 1950, **47**, 193–234.
14. Keller, F. S., & Schoenfeld, W. N. *Principles of psychology*. New York: Appleton-Century-Crofts, 1950.
15. Perkins, C. C., Jr., & Cacioppo, A. J. The effect of intermittent reinforcement on the change in extinction rate following successive reconditioning. *J. exp. Psychol.*, 1950, **40**, 794–801.
16. Postman, L. The history and present status of the law of effect. *Psychol. Bull.*, 1947, **44**, 489–563.
17. Sheffield, V. F. Extinction as a function of partial reinforcement and distribution of practice. *J. exp. Psychol.*, 1949, **39**, 511–525.
18. Skinner, B. F. *The behavior of organisms*. New York: Appleton-Century, 1938.
19. Thorndike, E. L. *Human learning*. New York: Century, 1931.
20. Thorndike, E. L. *The fundamentals of learning*. New York: Teachers College, Columbia Univ., 1932.
21. Thorndike, E. L. An experimental study of rewards. *Teach. Coll. Contr. Educ.*, No. 580, 1933.
22. Thorndike, E. L. *The psychology of wants, interests and attitudes*. New York: Appleton-Century, 1935.
23. Tolman, E. C. There is more than one kind of learning. *Psychol. Rev.* 1949, **56**, 144–155.
24. Witmer, L. R. The association value of three-place consonant syllables. *J. gen. Psychol.*, 1932, **7**, 80–103.

MOTIVATION IN LEARNING: XI. AN ANALYSIS OF ELECTRIC SHOCK FOR CORRECT RESPONSES INTO ITS AVOIDANCE AND ACCELERATING COMPONENTS [1]

KARL F. MUENZINGER, WILLIAM O. BROWN, WAYMAN J. CROW, AND ROBERT F. POWLOSKI

University of Colorado

In studies of electric shock for correct turns, it was found repeatedly that although learning of a black-white discrimination by white rats was definitely facilitated by shock for correct turns it was never quite as effective as shock for wrong turns (2, p. 181). This is so in a corrective situation where an animal is permitted to retrace the wrong alley and enter the right one. In a non-corrective situation, where an animal is not permitted to retrace after entering the wrong alley, the accelerating effect of shock for correct turns seemed slight indeed (3), if not absent, as claimed by George J. Wischner (4). In attempting to account for this difference between the effects of shock for correct turns in a corrective and a non-corrective situation Muenzinger and Powloski postulated "that shock after the point of choice whether for right or wrong turns in a visual discrimination box has an accelerating function in both situations, non-corrective as well as corrective, but in addition it also has an avoidance function which may under certain circumstances mask the accelerating effect. Shock after the point of choice makes the animal more sensitive to the cues to be discriminated, but it also creates a tendency in an animal to avoid the alley in which it has been shocked. If it should be possible to train an animal to adapt itself to, and thus overcome, the avoidance effect it might then exhibit the accelerating effect more clearly" (3, p. 123).

In order to test this hypothesis, 30 white rats received preliminary training with electric shock in a straightaway, 30 rats received such training without shock, and 30 rats received no preliminary training at all, before they all learned a black-white discrimination in a non-corrective situation. In each of the three groups of 30 *S*s, 10 learned the discrimination without shock (no-shock), 10 with shock for wrong turns (shock-wrong), and 10 with shock for correct turns (shock-right). According to the hypothesis the shock-right *S*s that had preliminary training with shock should now show a learning record approaching that of the shock-wrong *S*s. In order to answer the question whether this expected result would actually be due to the preliminary shock experience or merely to the preliminary training in the straightaway, the additional subgroup with such training, but without shock, was needed.

Conditions and Methods of Training

The apparatus, a T-shaped discrimination box, has been described in the preceding article (3). In addition, a straightaway was used, 15 in. long and 3 in. wide, with a starting box at one end and a goal box at the other. The floor of the middle section consisted of two solid

[1] The authors wish to thank the Council on Research and Creative Activity of the University of Colorado whose financial assistance made this study possible.

copper strips 15 in. long and slightly less than 1½ in. wide. They were about ½ in. apart at the center of the alley and formed a shallow trough so that any feces would drop through. They functioned as an electric grid.[2]

Ninety white rats of Sprague-Dawley strain were divided into nine groups as indicated in Table 1. General handling and feeding were also the same as in the preceding experiment (3). The rats were motivated by a 23-hr. food deprivation.

Six of the nine groups were given preliminary training in the straightaway for 12 days of ten runs each. At the end of each run *S* was rewarded with food in the goal box. The first four runs of the first day were used to accustom *S* to the apparatus. It was first placed in the goal box and then farther and farther away and finally with the fifth run *S* was placed in the starting box and allowed to proceed to the goal box.

From the second to the twelfth day three of the groups were run without change of conditions. They constitute the 30 "preliminary training without shock" *S*s which were treated differentially only in the discrimination box where one group of 10 *S*s received no shock at all (Group Id), 10 received shock for wrong turns (Group IId), and 10 for correct turns (Group IIId).

The other 30 *S*s, referred to as having "preliminary training with shock," received shock in the straightaway at every other trial beginning with the second trial on the second day, or the twelfth trial counting from the beginning. The strength of the electric current was gradually increased from 30 to 100 microamperes from Trial 12 to Trial 42, and then 20 microamperes per day reaching 240 on the last day. In the discrimination box, where the current was never above 100 microamperes, the 30 *S*s were also treated differentially, 10 receiving no shock (Group Ic), 10 shock for wrong (Group IIc), and 10 shock for correct turns (Group IIIc).

Three of the nine groups (Ib, IIb, and IIIb) were given no preliminary training in the straightaway. They were the same as the groups with similar labels in the preceding experiment (3).

The handling of *S*s in the discrimination box was exactly the same as described in the preceding article, including the two days of accustoming them to the apparatus. The lighted panel was again the positive cue, and the criterion of learning consisted of two successive errorless daily runs of ten trials each. The method was noncorrective, that is, an *S* having made a wrong turn was not permitted to retrace, but had to proceed forward to the starting point, from there again to the choice point, and so on.

No *E* handled more than five *S*s in any one group. The senior author was responsible for planning the experiment and writing this report.

TABLE 1

NON-CORRECTIVE LEARNING CONDITIONS
(N = 10 for each group)

Group	Pre-training in Straightaway	Shock in Discrimination Box
Ib*	none	none
IIb*	none	for wrong turn
IIIb*	none	for correct turn
Ic	with shock	none
IIc	with shock	for wrong turn
IIIc	with shock	for correct turn
Id	without shock	none
IId	without shock	for wrong turn
IIId	without shock	for correct turn

* Groups Ib, IIb, and IIIb are the same as those with similar labels in the preceding experiment (3).

[2] Credit for suggesting this simple and effective type of grid should go to Dr. Karl Zener of Duke University to whom the authors hereby express their thanks.

RESULTS

The results are stated in terms of the mean number of "reinforcements" and "wrong turns" to reach the criterion. This terminology, replacing the usual terms "trials" and "errors," was adopted for reasons given elsewhere (3).

Our major concern was the question whether the accelerating effect of electric shock for correct turns in the non-corrective situation would show itself clearly if *S* had by previous training learned to overcome the avoidance effect of shock. The results show that this is so. Without previous experience with shock the shock-right group is significantly poorer in terms of reinforcements and wrong turns than the shock-wrong group, as shown in comparing Groups IIIb and IIb. However, given such experience, the difference between the shock-right and shock-wrong groups becomes insignificant, as shown by comparing Groups IIIc and IIc.

TABLE 2

MEAN REINFORCEMENTS AND WRONG TURNS PER *S* TO THE CRITERION

Condition in Discrimination Box	No Pre-Training					Pre-Training with Shock					Pre-Training without Shock				
	Gp.	Reinf.		Wrong Turns		Gp.	Reinf.		Wrong Turns		Gp.	Reinf.		Wrong Turns	
		M	*SD*	*M*	*SD*		*M*	*SD*	*M*	*SD*		*M*	*SD*	*M*	*SD*
No-shock	Ib	156	50.0	41.9	12.2	Ic	195	80.5	61.2	29.6	Id	164	46.5	47.2	19.5
Shock-wrong	IIb	78	14.7	19.2	3.0	IIc	72	13.3	17.2	6.8	IId	69	13.0	13.3	5.1
Shock-right	IIIb	120	39.0	33.0	6.9	IIIc	82	28.2	22.0	7.8	IIId	114	27.6	33.2	7.2

Another way of demonstrating the accelerating effect of shock-right is by comparing the shock-right group which had experience with shock prior to the discrimination box, and the other shock-right groups that had not. In terms of both reinforcements and wrong turns, Group IIIc, having had pretraining with shock, is significantly better than Groups IIIb and IIId which had their first experience with shock in the discrimination box.

Since it seemed possible that pretraining alone in the straightaway might have produced these results, even if no shock had been present during this phase of the experiment, a set of *S*s was given pretraining without shock in the straightaway before they were run in the discrimination box. The results in terms of reinforcements and wrong turns show that the shock-right groups with and without such pretraining do not differ significantly, as seen in comparing Groups IIIb and IIId.

And again, with no previous shock experience, although with pretraining in the straightaway, the shock-right group is significantly poorer than the shock-wrong group, as seen in comparing Groups IIId and IId, whose records closely resemble those of Groups IIIb and IIb, the original shock-right and shock-wrong groups without any pretraining.

TABLE 3

p-VALUES OF DIFFERENCES BETWEEN MEANS OF BOTH REINFORCEMENTS AND WRONG TURNS *

Group	IIb	IIc	IId	IIIc	IIId
IIIb	<.01	>.05	<.01	<.01	>.05
IIIc					
IIId					

* These values were obtained by Festinger's method for skewed samples (1).

DISCUSSION

The close resemblance of the learning efficiency of a shock-right group in a non-corrective situation to that of a shock-wrong group when both have been given previous experience with shock as compared with the significant difference between similar groups without such experience can have only one meaning: shock in this learning situation has essentially two effects, avoidance and acceleration.

If an animal is shocked when it makes a correct turn this experience first tends to bring about an avoidance of shock. It is only after the tendency to avoid the shock has been somewhat overcome that the accelerating function of shock after the point of choice can manifest itself unhindered. If the first experience with shock occurs in the discrimination

box and the tendency to avoid it has to be overcome there, the course of learning by shock for correct turns will be slowed down as shown in the records of Wischner's *Ss* and of our Groups IIIb and IIId. If, however, the tendency to avoid shock has been overcome by experience with shock prior to running the discrimination box, the accelerating effect of shock shows itself more immediately, and learning by shock for correct turns, as in Group IIIc, appears to be practically as accelerating as shock for wrong turns.

It might be asked whether pretraining with shock has a similar result in a corrective situation. One could hardly expect, however, to obtain more of an accelerating effect of shock for correct turns where it is already clearly manifested. Without pretraining the shock-right method produces a learning efficiency somewhat poorer than the one produced by the shock-wrong method, but the differences are not always significant. In an extension of the present experiment, the details of which are not reported here, three shock-right groups were trained in a corrective situation, one with no pretraining, one with pretraining in the straightaway with shock, and one with pretraining without shock. The mean reinforcements were 72, 72, 70, and the mean wrong turns 17.0, 16.7, and 16.6 respectively. In other words, prior experience with shock had no effect in a situation where the accelerating effect is pretty much in evidence already.

It is not possible to tell whether in this experiment the length of the pretraining in the straightaway and the intensity of the electric current are optimum values. Perhaps shorter pretraining and a smaller current might have had the same effects.

In the straightaway shock was administered only in alternate trials in order to simulate somewhat the succeeding experience in the discrimination box where, in the beginning at least, *Ss* would not experience shock after each choice. Again it is not possible to tell whether shock on every trial in the straightaway might not have been more effective. To have given shock on every trial in the straightaway might have created an advantage for the shock-right condition since shock on every trial could have led *Ss* to expect food after shock. This might have produced a preference for the shock side which would be the correct side for the shock-right group but the incorrect side for the shock-wrong group. By feeding on every trial and shocking on alternate trials one would expect that no systematic preference would exist for the shock or no-shock side in either Group II or III.

It might not be amiss to point out that under the conditions used here the rats, even though at first disturbed by shock in the straightaway, overcame their avoidance reaction fairly well and were willing eventually to cross a grid which must have given them a considerable shock. The gradual increase in current, no less than the frequent exposure to it (55 times), was probably an important factor in overcoming undesirable emotional effects, if such effects existed at the beginning, which is probable.

Summary

The purpose of this study was to find out whether the accelerating effect of electric shock for correct choice (in a non-corrective situation) would manifest itself more readily if the avoidance effect of shock had been overcome previously.

White rats were given 120 trials in

a straightaway and received shock in 55 of the trials. They were subsequently trained by the non-corrective method to learn a black-white discrimination. The *Ss* which were shocked for correct turns were only slightly poorer than those shocked for wrong turns. The differences were statistically insignificant.

Among other *Ss* that had not received any prior experience with shock in the straightaway those trained with shock for correct turns were significantly poorer than those trained with shock for wrong turns.

It is assumed that electric shock in discrimination learning has two functions: (a) producing avoidance behavior, and (b) accelerating learning. If the tendency to avoid shock is overcome prior to training in the discrimination box, the accelerating function will manifest itself unequivocally.

(Manuscript received April 23, 1951)

References

1. Festinger, L. A statistical test for means of samples from skew populations. *Psychometrika*, 1943, **8**, 205–210.
2. Muenzinger, K. F., Bernstone, A. H., & Richards, L. Motivation in learning. VIII. Equal amounts of electric shock for right and wrong responses in a visual discrimination habit. *J. comp. Psychol.*, 1938, **26**, 177–186.
3. Muenzinger, K. F., & Powloski, R. F. Motivation in learning. X. Comparison of electric shock for correct turns in a corrective and a non-corrective situation. *J. exp. Psychol.*, 1951, **42**, 118–124.
4. Wischner, G. J. The effect of punishment on discrimination learning in a non-correction situation. *J. exp. Psychol.*, 1947, **37**, 271–284.

THE EFFECT OF DEGREE OF SHADING CONTRAST IN INK BLOTS ON VERBAL RESPONSE

JOHN C. BALLOCH
Fisk University[1]

Recent years have seen a tremendous growth in the popularity of the so-called projective techniques. Among the most prominent of these techniques is the Rorschach method of personality evaluation utilizing ten standardized ink blots (6).

As the Rorschach method has grown in importance as a clinical tool, the amount of importance attached to the shading responses as specific indicators of types of adjustment or maladjustment has also increased. By the term shading responses, we mean those verbal productions which are determined wholly or in part by the shading nuances which appear in practically all the blots. They are the responses to the stimuli along the black-white axis of the color pyramid, including responses to black, white, and intermediate greys.

Less is known about the significance of the shading responses than any other variable in the Rorschach. This is partly because the significance of shading was considered relatively late in the use of the Rorschach test, and partly because it is the one response which Rorschach himself did not fully describe. The evidence, such as we have, is based on clinical observation and quite a large amount of it is purely speculative. There is a paucity of experimental data in all phases of Rorschach development. Rorschach himself, and most of his followers, have freely admitted this lack.

There is an especial lack of information concerning the relationship between type of response and the variables in the stimulus ink blot which provoked it. Gibson (3) states that the Rorschach method ". . . presupposes a *non-relationship* between the stimulus pattern and the perception. . . ." However, as Gibson goes on to point out, such a relationship does exist. The present study is an attempt to shed some light on the nature of this relationship through varying one aspect of the stimulus ink blot, the degree of shading contrast.

Of the many stimulus variables present in an ink blot, the shading variable was chosen for study because of the importance which clinicians attach to its use as indicators of maladjustment. Beck (1) states that the use of shading to give vista effects indicates feeling of inferiority and depression while the use of shading to give hazy, diffuse effects, as clouds or smoke, indicates passive reaction to threat. Klopfer and Kelley (5) interpret the use of diffusion responses (*K*) as indicating feelings of insecurity and anxiety while the use of shading to give surface color to an object (*C'*) represents feelings of depression.

It was thought that a study of the shading contrast variable in the stimulus ink blot might (a) throw further light on the general problem of to what extent the ink blot influences the response of *S*; (b) more specifically, indicate to what extent the amount and kind of shading responses given

[1] The data contained in this article were gathered at Michigan State College as part of a doctoral thesis. The author wishes to express his gratitude to Dr. S. Howard Bartley for his help and guidance in directing this study.

by *S* are functions of this variable, and finally, (c) by isolating a factor which may influence the production of particular kinds of shading responses, point the way to the development of ink blots specifically designed to elicit these responses, much as Levy (7) has developed plates specifically designed to elicit human movement (*M*) responses. Such ink blots might have value in experimental designs where the shading response is the primary variable under consideration.

Method

Subjects.—The population for this experiment consisted of 75 students at Michigan State College. These *S*s were divided into three groups of 25 each, equated for age and sex. There were 18 women and 7 men in each group, and the mean ages of the groups were 20.4, 19.8, and 20.1 yr.

All *S*s were volunteers. A majority of the psychology majors used in the experiment were enrolled at the time in the beginning experimental psychology course at Michigan State College. The non-psychology major students came from a wide variety of fields.

Previous test experience with the Rorschach or projective techniques disqualified a student as an *S* for this experiment.

Apparatus and test materials.—A couch with reclining back was provided for *S*. The stimulus cards were held in a fixed position 60 cm. from *S*'s eyes by a suitable stand. A 60-w. lamp, 120 cm. from the card, provided illumination.

TABLE 1

Reflectance of Control Greys Following Photographic Reproduction

Number of Grey	Per Cent Reflectance (Incandescent Light)		
	Azo paper No. 0	Azo paper No. 2	Azo paper No. 4
1	45.79	45.57	50.02
2	43.64	44.11	48.73
3	42.79	43.61	47.31
4	40.38	42.28	41.08
5	36.85	35.96	31.66
6	31.19	27.84	21.42
7	25.28	23.76	9.39
8	18.14	11.03	1.58
9	14.50	6.92	1.14
10	11.91	4.01	1.04

TABLE 2

Experimental Design

Group	Shading Contrast		
	Weak	Medium	Strong
I	B*	C	A
II	A	B	C
III	C	A	B

* Refers to batches of seven cards.

The first seven cards of the Rorschach (6), Behn-Rorschach (8), and Harrower (4) series of ink blots were used as the stimuli. These 21 ink blots were photographed on Kodak Commercial film, a film which seemed to give the best rendition of the red coloring on Cards II and III in each series. Each of the resulting negatives was then reproduced by the contact process on Kodak Azo papers No. 0, 2, and 4, making a total of 63 ink blots, 21 of light shading contrast, 21 of medium contrast, and 21 of strong shading contrast. From data supplied by the manufacturer (2) it appeared that Azo papers No. 0, 2, and 4 would yield approximately equal intervals of shading contrast.

At the time the 21 ink blots were photographed, a control figure consisting of a series of greys was also photographed, and treated by the same photographic process. This control negative was also reproduced on Kodak Azo papers No. 0, 2, and 4. All the cards and the control greys were photographed, developed, and printed at the same time under similar conditions.

The reflectance of the control greys was then determined by the use of the Macbeth illuminometer. The results appear in Table 1.

Procedure.—Each of the 21 ink blots was assigned to one of three batches of seven cards. The three groups of *S*s were shown the stimulus cards in a modified Latin square design as shown in Table 2. An *S* in Group I, for example, saw a card which appeared in Batch A under strong conditions of shading. All *S*s saw all of the 21 ink blot forms, seven with weak shading contrast, seven with medium shading contrast, and seven with strong shading contrast. The order of presentation of the stimuli was random.

Since the presence of shading in ink blots is thought by Rorschach workers to be emotionally disturbing to some *S*s, continuous records of circulatory and respiratory changes were made. However, since no changes in physiological responses were found with respect to the degree of shading contrast, the present report will be confined to the verbal productions of *S*s.

The *S* was given the following instructions: "People see all sorts of things in these ink blot pictures, now tell me what you see, what it might be for you, what it makes you think of."[2]

The first card was placed in the card holder and all responses were recorded verbatim by *E*. When *S* indicated that he was through, the card was removed, the next card placed in the holder, and the procedure repeated. If *S* was not through at the end of 90 sec., he was allowed to finish the response he was giving after which the card was exchanged. This procedure continued until all 21 ink blots had been seen.

In general, scoring criteria according to Klopfer and Kelley (5, pp. 119–141) were used to determine into which shading category a response fell. Since there was no inquiry, only the performance proper or "free association" responses could be utilized in evaluating the type of shading used.

1. Depth responses (*FK*) were scored when the concept was a landscape, scene, valley, mountain, cave, etc. Reflections were not scored *FK* unless the reflected object itself utilized shading, i.e., "a lake shore with surrounding trees and brush reflected in the water" was scored *FK*, while "a person looking at his reflection in a mirror" was not.
2. Diffuse responses (*K*) were scored when the object seen was a cloud, fog, haze, or smoke.
3. Surface color (*C'*) was scored when the object seen was described as being black, white, or intermediate shades of grey or where the object was described as being striped.
4. Toned-down depth (*k*) was scored when the object seen was an X-ray or a topographical map.
5. Surface texture responses (*c*) were responses where the seen object was described as being furry, rough, grainy, hairy, pebbly, etc., or where highlights were specifically mentioned.
6. The number of times shading was used as a determinant on each card was found by adding the five shading categories, above, for each card.

The responses were scored without regard to form (whether predominant or subordinate) and no distinction was made between main and additional responses. Concept organization was not considered.

Results

The effect of degree of shading contrast on the number of times shading was used in responding to the ink blots was determined by an analysis of variance (Table 3). This analysis indicates that the number of times shading was used did not vary significantly with the degree of shading contrast.

Table 4, which summarizes the findings with respect to the various kinds of shading used, indicates that the number of times shading was used as surface texture (*c*) did not vary significantly with the shading contrast. The shading response categories, diffuse effects (*K*) and toned-down depth effects (*k*), likewise showed no significant increases or decreases with respect to this variable.

However, as the degree of shading contrast in the ink blots was increased, there was a significant increase in the number of times shading was used as achromatic color (*C'*). The total number of *C'* responses given under weak conditions of shading contrast was 47, under medium conditions of contrast it was 60, and under strong conditions it was 70. These differences are significant at the 5% level of confidence.

The number of responses utilizing shading as depth (*FK*) also varied significantly with the degree of shading contrast. Under weak conditions of shading, the total number of *FK* responses was 26, under medium condi-

TABLE 3

Effect of Degree of Shading Contrast on the Production of Responses Using Shading

Source of Variance	*df*	*ss*	*F*
Groups	2	34.032	3.06
Shading	2	24.413	2.19
Batches	2	103.841	9.32*
Cards	18	2016.984	20.12*
Error	38	211.570	
Total	62	2390.840	

* Significant at the 1% level of confidence.

[2] Standard Rorschach instructions according to Klopfer and Kelley (5, p. 32).

TABLE 4

ANALYSIS OF VARIANCE OF TYPES OF RESPONSES

Source of Variance	df	Surface Texture		Achromatic Color		Depth		Diffuse Effects		Toned-Down Depth Effects	
		ss	F	ss	F	ss	F	ss	F	ss	F
Groups	2	1.365	<1	9.238	2.83	10.889	3.05	6.889	3.72*	3.937	4.07*
Shading	2	2.317	<1	6.333	3.89*	11.555	3.24*	.793	<1	.127	<1
Batches	2	42.317	3.82*	3.523	1.08	.413	<1	.508	<1	2.112	4.36*
Cards	18	943.079	9.45**	440.381	15.02**	198.317	6.17**	91.651	5.51**	12.889	1.48
Error	38	210.668		61.906		67.810		35.143		18.380	
Total	62	1199.746		527.714		288.984		134.984		39.556	

* Significant at the 5% level of confidence.
** Significant at the 1% level of confidence.

tions it was 50, and under strong conditions it was 40. This finding is significant at the 5% level of confidence.

DISCUSSION

One finding of the present study is that increasing the degree of shading contrast results in an increase in the number of responses given which utilize the shading to give surface color to the thing seen (*C'*). A possible explanation for this increase may be found in an examination of what happens to the ink blot when the degree of shading contrast is increased photographically. When this is done, the number and amount of intermediate greys which appear in the ink blot are reduced; the darker greys tend toward the black end of the scale and the lighter greys tend toward white. Thus an ink blot reproduced under strong contrast conditions abounds in black and white whereas the same ink blot reproduced under weak contrast conditions contains hardly any black or white and abounds in intermediate greys. Black and white as surface colors may lend themselves to more objects in nature than do the intermediate greys, resulting in an increase in the number of responses utilizing surface color when more black or white is present.

Another explanation might be that the increased contrast between the black and white itself serves to attract the attention of *S* to the achromatic color, resulting in an increased use of the achromatic color in his response.

As a result of varying the degree of shading contrast in the ink blots there are also differences in the number of responses given utilizing the shading as depth. The difference here does not seem to be proportional to the degree of shading contrast in the cards; the ink blot with the weakest shading contrast produced fewest depth responses but the medium shaded cards produced more depth responses than did the cards with the strongest degree of shading contrast. The only significant difference is between the weak shading card and the medium shaded card.

An examination of the three cards which produced the most responses utilizing the shading to give depth to the response prompts three possible explanations. The process of increasing the contrast tends to lighten the greys which appear at the edges of the blots, causing them to blend with the white surround. The resulting haziness of outline gives an aerial perspective to the blot forms, resulting in more seen depth as the contrast is

increased. Secondly, the increased number of depth responses may be attributed to the secondary cue overlay. Some of the ink blots seem to have layers superposed one on another. With increased contrast between the black and white, the overlaying layers may have become more apparent to *S*. Thirdly, the increased contrast may enhance the depth producing attached shadow effect of some parts of the blots.

Summary

Seventy-five *S*s were shown ink blots reproduced under weak, medium, and strong conditions of shading contrast. Responses utilizing the shading to give surface texture (*c*), diffusion effects (*K*), and toned-down depth effects (*k*) did not vary with contrast, nor were there changes in the number of shading responses produced. However, it was found that there was an increased tendency to produce depth responses (*FK*) and to utilize the shading as surface color (*C'*) when the shading contrast was increased.

(Manuscript received April 24, 1951)

References

1. Beck, S. J. *Rorschach's test.* New York: Grune and Stratton, 1944.
2. Eastman Kodak Co. *Kodak data book—Kodak papers* (4th ed.). Rochester: Eastman Kodak Co., 1947.
3. Gibson, J. J. Studying perceptual phenomena. In T. G. Andrews (Ed.), *Methods of psychology.* New York: Wiley, 1948.
4. Harrower, M. S. *Psychodiagnostic inkblots.* New York: Grune and Stratton, 1945.
5. Klopfer, B., & Kelley, D. M. *The Rorschach technique.* Yonkers-on-Hudson: World Book Co., 1942.
6. Rorschach, H. *Psychodiagnostics.* New York: Grune and Stratton, 1942.
7. Zubin, J., & Young, K. M. *Manual of projective and cognate techniques* (privately published, 1948).
8. Zulliger, H. *Sinfuhrung in den Behn-Rorschach Test.* Bern: Hans Huber, 1941.

RATE OF MANIPULATIVE LEARNING AS A FUNCTION OF GOAL-SETTING TECHNIQUES [1]

GERALD C. HELMSTADTER AND DOUGLAS S. ELLIS

Iowa State College

Attempts to maximize the motivation of individuals performing a task frequently involve the use of both knowledge of results and a goal against which *S*'s performance is compared. Although knowledge of results is a motivator of demonstrated significance (6), few systematic studies have investigated the combined motivational effects of knowledge of results and goal-setting manipulations.

The studies of Mace (5) and Bayton (1) are typical. Mace found that a moving standard based on previous performance was more effective in a computational task than simple instructions to "do your best." Both of these techniques were superior to a fixed goal set by *E*. The least effective technique was simple instructions to surpass previous performance.

Working with a simple manipulative task, Bayton found that four motivational techniques could be ranked from most effective to least effective as follows: (1) *S* sets own goal and tells *E* what it is; (2) *S* sets own goal but does not tell *E* what it is; (3) *S* is given knowledge of results with no other comment; (4) *S* is told nothing concerning his previous performance. Statistical analysis revealed that the only significant differences were between Methods 1 and 4 and between Methods 2 and 4.

These two studies suggest at least three parameters that may define the influence of motivational techniques which combine knowledge of results with a standard or goal. Mace's finding that a moving standard was more effective than a fixed one indicates that the relationship between the standard and *S*'s performance is important for externally set goals. Bayton's results suggest that, for self-set goals, the degree of *S*'s ego involvement in his goal should be considered. Finally, whether the goal is self-set (Bayton's most effective manipulation) or externally-set (Mace's most effective manipulation) may be a critical dimension of goal-setting techniques.

The present study explores this last dimension by utilizing the following four motivational conditions:

I. Knowledge of results.
II. Self-set: *S* sets own goal on the basis of past performance.[2]
III. Externally-set norm: *E* provides a goal based on an average performance curve.
IV. Externally-set improvement: *E* provides a goal based on *S*'s previous performance and increments on an average performance curve.

Conditions II and IV are, respectively, the most effective techniques used by Bayton and Mace. Their comparison should provide evidence of the importance of who sets the goal. Condition III, a technique commonly used in practical situations, serves as an additional externally-set goal condition for comparison with the self-set goal of Condition II. Condition I is included as a control to determine if any of the goal-setting techniques used provide motivation beyond that associated with simple knowledge of results.

METHOD

Subjects and experimental design.—A total of 100 volunteers from undergraduate psychology courses at Iowa State College served as *S*s. Except for the first five *S*s who were placed in Group I, they were assigned to the four groups corresponding to the four motivational conditions so that the groups were fairly well matched on

[1] Paper presented at the 1951 meeting of the Iowa Academy of Science.

[2] This method of goal setting is equivalent to the manipulation followed in level of aspiration experiments (7).

initial score. The sex distribution for the groups was as follows: Groups I, III, and IV: 12 females, 13 males; Group II: 11 females, 14 males.

Task.—The task used was a modified form of the block-turning portion of the *Minnesota Rate of Manipulation Test.* The *S* was required to work continuously through the middle two rows of blocks, and was given standardized instructions and a demonstration of the work methods to be used. Following an untimed practice trial, ten 60-sec. trials were administered with a 60-sec. rest interpolated between trials. The number of blocks turned per 60-sec. trial was taken as the measure of performance.

Goal-setting procedures.—During rest periods, knowledge of results was given to all *S*s by showing *S* a consecutive plot of his previous performance scores. The *S*s in the knowledge of results group (Group I) received no additional treatment. The *S*s in the self-set goal group (Group II) were asked to state an expected score for their next trial, and this aspiration score was also plotted on the graph. A complete average performance or norm curve against which *S* could compare his performance was shown to members of the externally-set norm goal group (Group III) after each trial. For members of the externally-set improvement goal group (Group IV), the goal set for *S* on each trial was determined by adding to his previous performance the increment on the norm curve between the same two trials.[3]

Results

The learning curves of the four groups are presented in Fig. 1. The apparent similarity of the curves is

[3] The norm curve was an estimate based on the results of previous investigations (3, 4) as well as the scores of the first five *S*s tested in the present experiment. Evidence of the reasonableness of the various goals used is presented in Fig. 2.

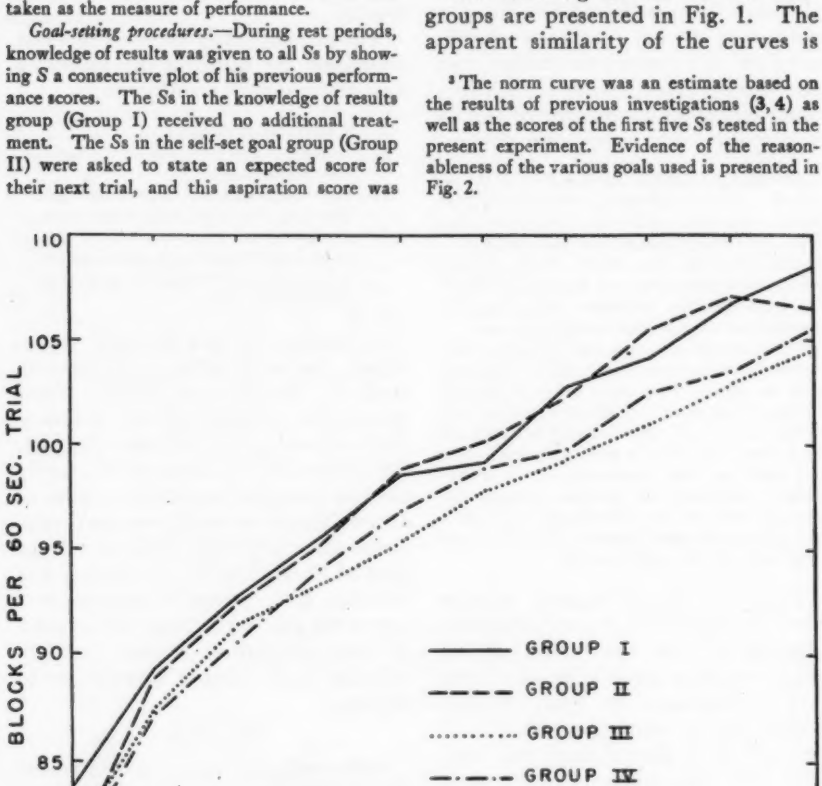

Fig. 1. Block-turning performance as a function of motivational techniques. Group I—knowledge of results; Group II—self-set goal; Group III—externally-set norm goal; Group IV—externally-set improvement goal.

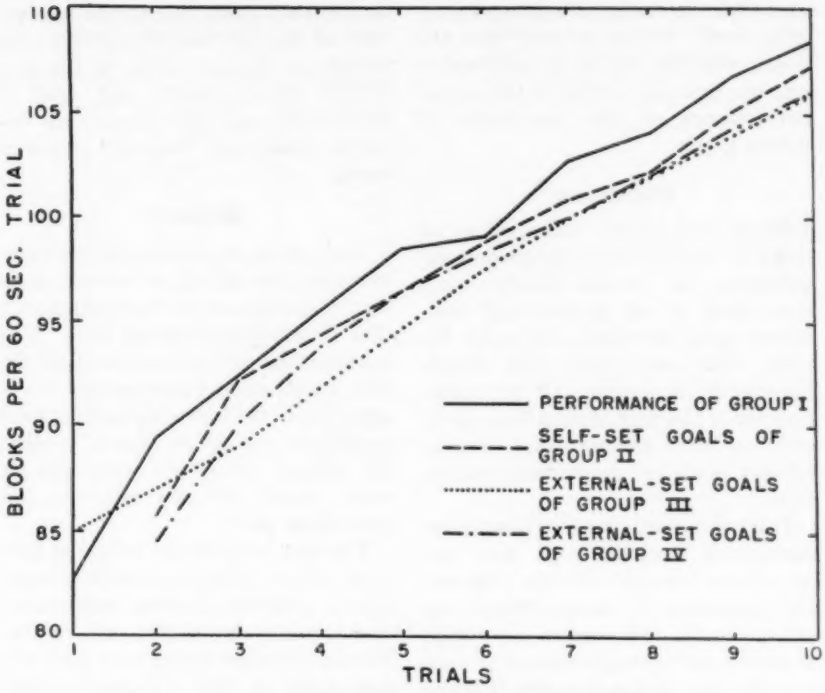

FIG. 2. Goals associated with the three goal-setting techniques and performance of the knowledge of results group (Group I).

supported by the analysis of variance arrayed in Table 1,[4] which indicates that none of the goal-setting techniques differed significantly in performance improvement (Trial 10 minus Trial 1) during the course of the experiment. The significant sex difference indicates that the males showed greater improvement than the females.[5]

Figure 2 compares the average performance of the knowledge of results group (Group I) with the goals associ-

TABLE 1

ANALYSIS OF VARIANCE OF IMPROVEMENT SCORES

Source of Variation	*df*	*ss*	*F*
Goal-setting techniques (GST)	3	84.55	.41
Sex	1	275.61	4.01*
Sex × GST	3	103.04	.50
Error	92	6320.91	
Total	99	6784.11	

* Significant at the 5% level of confidence. The 5% and 1% values of *F* are, respectively, 3.95 and 6.92.

[4] This analysis was based on Snedecor's method for disproportionate subclass numbers (8, pp. 289–290). Appropriate statistical tests (2, p. 11; 8, p. 285) indicated that the data met the assumptions necessary for such an analysis.

[5] Although the use of difference scores minimizes differences in initial ability of the groups, an analysis of variance of initial scores showed no statistically significant difference between the groups or sexes in respect to initial ability. The fact that the sexes differed in improvement but not in initial ability might be due to such factors as (a) differences in motivation, (b) differences in amount of pre-experimental practice on similar tasks, (c) sampling errors, etc.

ated with the three goal-setting techniques used. It can be seen that the goals, whether self-set or externally-set, are generally close to the actual performance of the knowledge of results group.

Discussion

Since performance was used as an index of motivation in the present experiment, the results clearly imply that none of the goal-setting techniques used provided motivation beyond that associated with simple knowledge of results. Of particular interest is the fact that self-set goals were no more effective than externally-set goals of comparable magnitude.

It is, of course, possible that the experimental situation used may not have been favorable for the differential operation of externally-set and self-set goals. The use of knowledge of results with college student *Ss* may provide such high motivation that the effects of additional motivational manipulations are negligible. The simplicity of the task and the use of distributed practice may have operated to reduce the level of motivation sufficient for maximum performance, which would also tend to decrease differences between experimental conditions. Finally, the closeness of the externally-set and self-set goals to actual performance and to each other should be mentioned. It is possible that any differential effect of externally-set and self-set goals would depend on the goals being set quite a distance above actual performance.

These considerations suggest the desirability of examining the following variables in further investigations of the relative influence of externally-set and self-set goals: (a) initial motivation of *Ss*; (b) task complexity; (c) inhibitory characteristics of the conditions under which the task is performed, and (d) relationship between goals and "normal" performance.

Summary

An experiment was conducted to investigate the effects of several goal-setting techniques on the performance of a simple block-turning task. The investigation utilized a total of 100 *Ss*, distributed among four groups corresponding to the following motivational conditions: (a) knowledge of results; (b) self-set goal; (c) externally-set norm goal; (d) externally-set improvement goal.

The obtained results indicated that none of the three goal-setting techniques provided greater motivation than simple knowledge of results. Several variables which may limit the generality of this conclusion were specified.

(Manuscript received May 4, 1951)

References

1. Bayton, J. A. Performance as a function of expressed and non-expressed levels of aspiration. *Amer. Psychologist*, 1948, **3**, 274. (Abstract.)
2. Bowles, R. L. Accuracy of certain approximate methods in predicting the correct model for experiments with unequal frequencies in the subclass. Unpublished M.A. thesis, Iowa State College, 1950.
3. Ellis, D. S. Speed of manipulative performance as a function of work-surface height. *J. appl. Psychol*, 1951, **35**, 289–296.
4. Kimble, G. A., & Bilodeau, E. A. Work and rest as variables in cyclical motor learning. *J. exp. Psychol.*, 1949, **39**, 150–157.
5. Mace, C. A. Incentivies: Some experi-

mental studies. *Ind. Health Research Board.* (Great Britain), 1935, Report No. 72.

6. Manzer, G. W. The effect of knowledge of output on muscular work. *J. exp. Psychol.*, 1935, **18**, 80–90.

7. Rotter, J. B. Level of aspiration as a method of studying personality: I. A critical review of methodology. *Psychol. Rev.*, 1942, **49**, 463–474.

8. Snedecor, G. W. *Statistical methods.* Ames: Iowa State College Press, 1946.

REACTIVE INHIBITION AS A FACTOR IN MAZE LEARNING: III. EFFECTS IN THE HUMAN STYLUS MAZE [1]

MERRELL E. THOMPSON

University of Arkansas

This study is the third in a series of investigations of the role of reactive inhibition in maze learning. In the first study (7), Hull's assumption that reactive inhibition is an increasing function of work was confirmed in a maze situation. In the second study (8) the role of reactive inhibition was demonstrated in studies of place learning versus response learning. The purpose of the present study is to extend the application of the principle of reactive inhibition to encompass the aspect of maze behavior formerly attributed to centrifugal swing.

Hull (5) has assumed that each reaction of an organism builds up a state of negative motivation within the organism which tends to prevent the repetition of that response for a short time. This state of negative motivation is temporary in character, most of it showing spontaneous dissipation after a period of time. This means that when an animal makes a choice in a maze (e.g., the right alley) and receives an appropriate reward, two things occur: (a) the act of turning right is strengthened, and (b) inhibition against turning right again for some time is developed. If the next trial follows immediately, the animal will tend to alternate his response, that is, turn left at the choice point.

This is somewhat similar to the maze situation in which the animal makes a forced turn preceding the choice point and then tends to choose the alley in the opposite direction. This latter phenomenon has been formerly interpreted in terms of centrifugal swing which was originally derived from Schneirla's studies on learning and orientation in ants (6) and denotes a relationship between turns at a choice point and the immediately preceding pattern of movement forced by the spatial arrangement of the maze alleys. Later, Ballachey and Krechevsky (2) and Ballachey and Buel (1) demonstrated the operation of centrifugal swing in the choice-point behavior of rats. Still more recently, Bitterman and Bretz (3) have applied the concept to human maze learning. The basic assumption of this hypothesis is that an organism tends to be forced against the wall away from the direction of a forced turn, thereby facilitating a choice in the direction opposite to that of the preceding forced turn. Thus, we would expect a preponderance of left turns at a choice point which was just preceded by a forced right turn.

While the experimental evidence on this point leaves little to be desired, the explanation of the phenomenon, from a behavioral point of view, is far from adequate. The chief weakness of this hypothesis is that there is no way of incorporating it into a broader systematic theory of learning. The hypothesis of centrifugal swing may provide an adequate vehicle for the explanation of choice-point behavior following a forced turn but it adds nothing to our understanding of other maze behavior.

The principle of reactive inhibition, on the other hand, has the advantage of explaining a much wider range of phenomena. Hull has successfully employed it in his explanations of many of the phenomena of conditioning and mention has been made of two previous attempts to extend the principle to encompass certain aspects of maze behavior. The present study is an attempt to show that the principle of reactive inhibition may more adequately cover maze behavior formerly attributed to centrifugal swing.

The prediction of choice-point behavior in all the maze patterns characterized by one preceding forced turn would be the same whether one held to the hypothesis of centrifugal

[1] This paper is a modified version of a report presented by the writer at the 1950 meeting of the APA. The writer wishes to express his gratitude to Dr. R. H. Burros for his advice on the statistical treatment of the data.

swing or to that of reactive inhibition. However, since centrifugal swing, due to its physical character, cannot accumulate (it is entirely dependent upon the just preceding turn) it is possible to arrange a maze situation which critically tests the two assumptions.

If we have *two or more* forced consecutive like turns preceding the choice point, we will expect, if centrifugal swing is the determining factor, no greater proportion of opposing turns at the choice point than in a maze where there is only one preceding forced turn. However, with the principle of reactive inhibition, which assumes that the inhibition is a cumulative affair and that two like turns will develop more inhibition than one turn, we will expect more opposing turns at the choice point when it is preceded by two forced like turns than by only one forced turn.

Experimental Procedure

Subjects.—The *S*s were 120 college students from the general and experimental psychology courses at Kent State University. There were 79 males and 41 females assigned at random to the various groups.

Apparatus.—The present investigation utilized single-choice, grooved-type stylus mazes. In Fig. 1 are illustrated the three mazes, each in four positions, used in the study. It is to be noted that Maze 1 consisted of a choice point preceded by one clockwise turn, Maze 2 of a choice point preceded by two clockwise turns, and Maze 3 of a choice point preceded by three clockwise turns.

All three mazes were identical with respect to width of alleys, length of path to choice point, and length of alleys from choice point to the goals. They differed in only one respect, the number of clockwise turns preceding the point of choice. The alleys were ¼ in. deep and ¼ in. wide, the distance from starting point to choice point was 8 in., and the length of each alley from choice point to goal was 1 in. Each maze was fastened to a board 9 in. square.

Procedure.—The *S*s were assigned at random to one of the 12 maze patterns (three mazes in four positions) and given one trial on that pattern. The maze was placed in a grooved slot on a table behind a black cloth screen containing a narrow slit through which *S*'s arm was permitted to pass. When *S* arrived for the experimental session the following directions were read by *E*:

"We want to determine how *S*s will respond to different mazes. There are no correct or incorrect turns. I will place the stylus which you hold at the starting point; then say 'begin.' Start pushing the stylus down the maze alley. Do not hesitate at the turns but proceed until I say 'goal.' You are to hold the stylus in your preferred hand as you hold a pencil, like this (demonstrate). Keep your hand and arm off the table and maze. Are there any questions? If not, we will proceed."

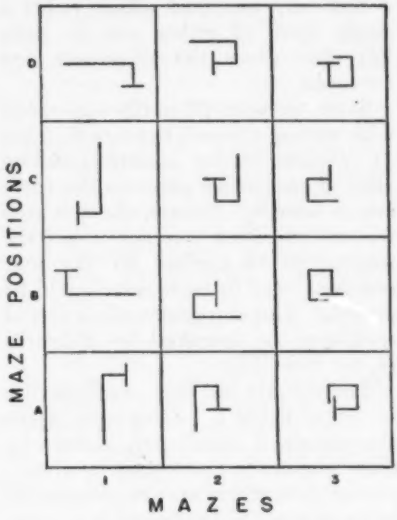

Fig. 1. The three mazes, each in four positions, used in the study

Results and Discussion

The design of the experiment permits an analysis in terms of the number of counterclockwise choices as a function of the position of the respective mazes and as a function of the number of preceding forced clockwise turns. Ordinarily this factorial design would permit a straightforward analysis of variance treatment. In the present study, however, such a treatment is not practical since each

S had only one trial which yields a single score of either one or zero. For this reason the chi-square test was used.

Since we were primarily concerned with testing the null hypothesis (that *Ss*' choices of the counterclockwise alley at the choice point in the three mazes were by chance), the first step was to determine whether or not the frequencies of choices for the four positions used in each maze could be pooled. This was essentially a test of technique as described by Edwards (**4**, pp. 108–113).

The results of this analysis are shown in Table 1. Since none of the chi-squares is significant, the variations within the maze position groups for the three mazes may be assumed to be no greater than expected in random sampling from a common population. This finding allows us to pool the data from the four subgroups for each maze and then to apply the chi-square test to the pooled frequencies. The hypothesis to be tested is that the three maze groups have been randomly drawn from a common population. The pooled frequencies of counterclockwise turns for Mazes 1, 2, and 3 were 16, 32, and 34, respectively. Since the obtained value of χ^2 (22.54) is highly significant for 2 *df* the null hypothesis tested must be regarded as untenable.

TABLE 1

NUMBER OF COUNTERCLOCKWISE CHOICES (N = 10 each group)

Position	Maze 1	Maze 2	Maze 3
A	7	6	7
B	4	9	8
C	3	9	10
D	2	8	9
χ^2	5.83	3.75	3.92
p	.12	.29	.27

The results of the present study clearly demonstrate that the choice-point behavior of human *Ss* is influenced by the number of preceding forced turns in the types of mazes employed. The effect of increasing the number of forced clockwise turns resulted in a significantly greater number of counterclockwise turns at the point of choice. In the present study the mechanics of movement through the alleys will not account for the relationship since the tendency for the counterclockwise choices increased as the number of preceding clockwise turns was increased.

If centrifugal swing is the predominant factor in choice-point behavior in these mazes we would expect the number of counterclockwise choices to be the same in all three choice-point situations regardless of the number of preceding clockwise turns. Unfortunately this aspect has been overlooked in the many attempts to explain choice-point behavior following forced turns in terms of centrifugal swing. The data from studies with rats (**9**) offered in support of this hypothesis have in fact shown the same relationship between the number of forced like turns preceding the choice point and choice-point behavior, although the implications of this relationship have not been noted.

In conclusion, we can say that, for human *Ss*, the number of counterclockwise choices will increase significantly as the number of forced clockwise turns is increased, the number of counterclockwise choices being some increasing function of the number of preceding clockwise turns. And we may add, in the interests of an integrated and systematic science, that since the principle of reactive inhibition can be applied to a broader range of behavioral phenomena, its use in explaining the effect of preced-

ing forced turns upon choice-point behavior is preferred over the use of the concept of centrifugal swing, limited as it is.

Summary

1. In order to determine which of two hypotheses, centrifugal swing or reactive inhibition, is more adequate in explaining choice-point behavior in maze patterns characterized by one or more forced like turns preceding the choice, 120 college students were given one trial each on one of 12 maze patterns. There were three single-choice, grooved-type stylus mazes (each in four positions) differing only in the number of forced clockwise turns preceding the choice point.

2. The over-all effect of increasing the number of forced clockwise turns is an increasing tendency for counter-clockwise alleys to be chosen at the point of choice as shown by a chi-square of 22.54 with 2 *df*.

3. It is suggested that the principle of reactive inhibition provides a more adequate explanation of this aspect of maze behavior than does the hypothesis of centrifugal swing.

(Manuscript received May 10, 1951)

References

1. Ballachey, E. L., & Buel, J. Centrifugal swing as a determinant of choice point behavior in the maze running of the white rat. *J. comp. Psychol.*, 1934, **17**, 201–223.
2. Ballachey, E. L., & Krechevsky, I. "Specific" versus "general" orientation factors in maze running. *Univ. Calif. Publ. Psychol.*, 1932, **6**, 83–97.
3. Bitterman, M. E., & Bretz, E. 'Centrifugal swing' effects in the human stylus maze. *Amer. J. Psychol.*, 1946, **59**, 267–272.
4. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
5. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
6. Schneirla, T. C. Learning and orientation in ants. *Comp. psychol. Monogr.*, 1929, **6**, 1–143.
7. Thompson, M. E. Reactive inhibition as a factor in maze learning: I. The work variable. *J. exp. Psychol.*, 1950, **40**, 131–134.
8. Thompson, M. E., & Thompson, J. P. Reactive inhibition as a factor in maze learning: II. The role of reactive inhibition in studies of place learning versus response learning. *J. exp. Psychol.*, 1949, **39**, 883–891.
9. Witkin, H. A., & Schneirla, T. C. Initial maze behavior as a function of maze design. *J. comp. Psychol.*, 1937, **23**, 275–304.

LATENT LEARNING IN A WATER MAZE

A. W. BENDIG[1]
University of Pittsburgh

In recent reports on studies of latent learning two aspects of the problem have been ascribed an increasing degree of importance: (a) the level of motivation of the *Ss* during the period when latent learning could be expected to occur, and (b) the ability of the *Ss* to discriminate the drive states used. Munn (7) points out that experiments involving weak motivation tend to demonstrate latent learning, whereas strong motivation tends to discourage such learning. The results of Meehl and MacCorquodale (5) confirm that rats strongly motivated tend not to show cognitive types of learning, but animals weakly motivated do. Such results suggest that Tolman's concept of the role of strongly motivating conditions in "narrowing the cognitive map" (9) may be applicable here.

As to motivational discriminability, Tolman (10) suggests that one of the important types of learning is that of "drive discrimination," and that these discriminations are particularly important in latent learning experiments. Heron's work on the ability of rats to discriminate different drives or different levels of the same drive (2) is relevant to this point. In discriminating between alternating states of hunger and thirst with one of the pair of drives satiated the animal may not be learning the difference between hunger and thirst drive stimuli, but simply the presence or absence of one of the sets of stimuli. Since these two drives are notoriously difficult to separate experimentally (2), the present research is concerned with using a different combination of basic drives: food deprivation and escape from water.

[1] These experiments were performed in the Laboratory of Neurophysiology, Western Psychiatric Institute and Clinic, and were supported by a research grant from the U. S. Public Health Service. Appreciation is expressed to Dr. Robert A. Patton for his advice and supervision during this research.

Experiment I

Method

Subjects.—Twenty-three male rats of the Sprague-Dawley strain were used. The *Ss* ranged in age from 130 to 150 days at the beginning of their latent learning training and were experimentally naive.

Apparatus.—The water-maze previously used in this laboratory (8) was modified to provide a T-maze pattern similar to that used by Kendler (4). The diagram of the maze is given in Fig. 1. The alleys were filled with water to a depth of 8½ in. and the water was maintained at a temperature of 60° F. Sliding retrace doors were placed at points *A* and *B* in Fig. 1. Metal landing platforms were placed at *X* and *Y* and were separated by a solid metal partition impossible for the *Ss* to scale.

Procedure.—Each *S* was given four trials per day during three days of training in the maze. On the first and third trials each day the *Ss* were placed in the maze at point *S* in Fig. 1, facing away from the choice point, and allowed to choose a right or left turn at the choice point. After reaching landing platform *X* or *Y*, *S* was allowed to remain for 30 sec. before being removed for the next trial. On the second and fourth trials each day *S* was forced to swim to the landing platform opposite to its previous choice. Forcing was done by closing the sliding

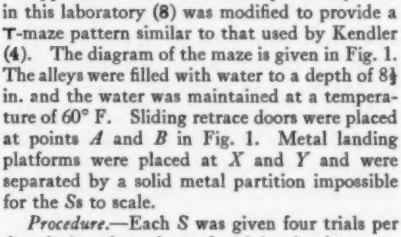

Fig. 1. Maze diagram

retrace doors at *A* or *B*. Thus, if *S* chose to turn to the right on his first trial and swim to *Y*, on his second trial he found door *B* closed and was forced to swim to *X*.

During the three training days, food (Purina Laboratory Chow) and water were always available in the *Ss*' home cages. In the maze, food was always found by the rats on one of the landing platforms and never on the other platform. Half of the *Ss* found food always at *X* and never at *Y*, and the reverse conditions obtained for the other half. The landing platform containing food had Purina dog pellets piled so that the *Ss* had to climb over and sit on the food during trials to that side. During the three training days, when the *Ss* always had food available in the home cages, none was observed to eat the food during the training trials.

After the last trial on the third training day the *Ss* were returned to the home cages from which all food had been removed. Twenty-three hours later they were given a single trial in the maze under the combined escape from water and (for the first time) hunger drives. The *Ss* were scored on this single trial as to whether they turned toward the side where they had always previously found food on the landing platform (food-side response) or toward the side where they had never found food (non-food-side response).

Results

On the single trial on the fourth day of the experiment, 14 of 23 *Ss* turned toward the side where they had always previously found food and 9 *Ss* made non-food-side response. Testing this distribution against a chance or equal probability hypothesis gives a chi-square value of 1.087 which with 1 *df* is not significant at the 5% level. The conclusion was drawn that, while a majority of the *Ss* performed as predicted by a latent learning hypothesis, the results could be parsimoniously attributed to chance.

Observations of *Ss* during the training and testing days suggested that one complicating factor was the strong emotional reaction of *Ss* to the water-maze situation. Such strong emotional response may have tended to interfere with cognitive learning of the maze or with the operation of the hunger drive on the fourth day. Two additional experimental groups were therefore run to determine the efficiency of certain changes in the experimental procedure.

Experiment II

Method

Subjects.—The same 23 Sprague-Dawley rats of the previous study were utilized again. The animals were 185 to 210 days old at the beginning of this study.

Apparatus.—The same maze was used as in the previous study.

Procedure.—The procedure was similar to that used in the previous study, but with certain changes. These changes were directed toward reducing the strength of emotional motivation during the training days of the experiment. (a) The training trials were preceded by five days of adaptation in a water-filled straightaway as used by Russell (8). Four trials per day were given and on each trial the *Ss* were required to swim a 48-in. straightaway to a landing platform. Adaptation to the swimming situation appeared to take place over the 20 trials in the straightaway. (b) The *Ss* were given four days of maze training while satiated for food instead of three days as in the previous study. (c) The *Ss* were placed in the maze at point *S* in Fig. 1 *facing toward* the choice point instead of facing away from the choice point. It was noted in the previous study that the *Ss* in making a 180° turn to approach the choice point tended to acquire a clockwise or counterclockwise turning habit. Such a habit may have tended to influence their response at the choice point. (d) Sliding retrace doors were placed at points *C* and *D* in Fig. 1 rather than at *A* and *B*. This tended to eliminate useless responses at the choice point. (e) The water temperature was kept at 77° F. instead of the previously used colder temperature. Using Wever's data (11) this should have tended to lower the strength of the escape-from-water drive. In all other respects the procedure followed was that used in Exp. I.

Results

On the single hunger-motivated trial on the fifth day, 17 of 23 *Ss* turned toward the food-side and 6 toward the non-food-side. Testing these proportions against a chance hypothesis gave a chi-square value of 5.260, which is significant at the 5%

level. These results suggested that the changes made in the training procedure increased the amount of latent learning shown by the *Ss*. To confirm this, an independent replication of the above experiment was performed as part of a study reported elsewhere (**1**).

Experiment III

Method

Subjects.—The *Ss* were 23 male albino rats of the strain maintained by Albino Farms. They were 116 to 128 days old at the beginning of the experiment.

Apparatus.—The same maze was used as in Exp. I and II.

Procedure.—The same procedure was used as in Exp. II with the single exception that all *Ss* were given four trials in the maze on the fifth day, the conditions of each of the four trials being the same as the single trial on Day 5 in Exp. II. Only the first trial on that day is relevant here.

Results

Eighteen of the 23 *Ss* turned toward the food-side on the first trial of the hunger-motivated fifth day. The chi-square value is 7.348, which is significant at the 1% level.

Discussion

The results suggest that the phenomenon of "latent learning" was present in Exp. II and III, if not in Exp. I. The *Ss* learned the location of food in the water maze when the food itself was not a reward object or (more precisely) was not a relevant reward for the dominant drive. However, the question can be raised as to whether the latent learning shown on the fifth day was actually "latent" during the training period. Meehl and MacCorquodale (**5**) note the appearance of an increasing preference for the food-side during the training trials. From the free-choice trials of the four training days (Trials 1, 3, 5, 7, 9, 11, 13, and 15) the percentage of free responses directed each day toward the food-side was computed for the separate groups of Exp. II and III and is found in Fig. 2. It is evident that a food-side preference was set up in the *Ss* during the training days. The somewhat different shapes of the curves of acquisition for Exp. II and III probably reflect sampling fluctuations, but may reflect strain differences. Whether this learning of the location of food and of showing a preference for the location when not experimentally motivated for the food is a function of Hull's "secondary reinforcement" (**3**), Miller's "acquired drives" (**6**), or Tolman's "equivalence beliefs" (**10**) cannot be answered by these experiments. However, some construct over and above that of primary reinforcement is a necessity in explaining latent learning.

It should be noted that latent learning was more evident in the groups of *Ss* that had been habituated to the maze and whose escape-from-water drive had been somewhat reduced by raising the temperature of the water. This is in accord with the results of Meehl and MacCorquodale (**5**) who found latent learning only when weak

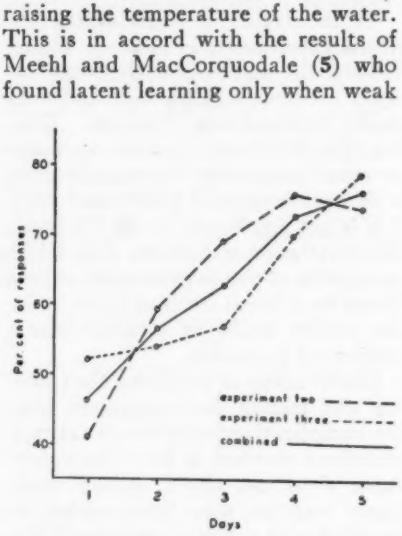

Fig. 2. Percentages of food-side responses on free-choice responses, by days

drives were present during the training sequence. It also agrees with Tolman's view of the effect of strong motivation (especially strong emotional motivation) on a narrowing of the cognitive map (9). In these more complex types of learning, strong motivation appears to have a disorganizing effect and the optimum drive must be weak for such learning to take place effectively.

Summary and Conclusions

1. The phenomenon of latent learning was demonstrated in a water maze. Escape-from-water and food-deprivation were used as motivating conditions in three experiments with rats, the training trials involving escape-from-water with food present and the *Ss* food-satiated, the test trial involving hunger motivation.

2. The amount of latent learning was a function of the amount of pretraining received by the *Ss*, and the influence of the pretraining was to reduce the unorganized drive in the *Ss*.

3. The latent learning is not "latent," but appears in performance during the training period.

(Manuscript received May 14, 1951)

References

1. Bendig, A. W. The effects of electroshock convulsions on latent learning in the white rat. Unpublished Ph.D. thesis, Univ. of Pittsburgh, 1951.
2. Heron, W. T. Internal stimuli and learning. *J. comp. physiol. Psychol.*, 1950, **42**, 486–492.
3. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
4. Kendler, H. H. An investigation of latent learning in a T-maze. *J. exp. Psychol.*, 1946, **36**, 212–220.
5. Meehl, P. E., & MacCorquodale, K. A further study of latent learning in the T-maze. *J. comp. physiol. Psychol.*, 1948, **41**, 372–396.
6. Miller, N. E. Studies of fear as an acquirable drive: I. Fear as motivation and fear-reduction as reinforcement in the learning of new responses. *J. exp. Psychol.*, 1948, **38**, 89–101.
7. Munn, N. L. *Handbook of psychological research on the rat.* New York: Houghton Mifflin, 1950.
8. Russell, R. W. Effect of electroshock convulsions on learning and retention in rats as functions of the difficulty of the task. *J. comp. physiol. Psychol.*, 1949, **42**, 137–142.
9. Tolman, E. C. Cognitive maps in rats and men. *Psychol. Rev.*, 1948, **55**, 189–208.
10. Tolman, E. C. There is more than one kind of learning. *Psychol. Rev.*, 1949, **56**, 144-155.
11. Wever, E. G. Water temperature as an incentive to swimming activity in the rat. *J. comp. Psychol.*, 1932, **14**, 219–224.

STUDIES IN SHORT-DURATION AUDITORY FATIGUE: II. RECOVERY TIME

ANITA I. RAWNSLEY AND J. DONALD HARRIS

U. S. Naval Medical Research Laboratory

The use of short duration stimulation is a very convenient approach to the many problems raised in the area of auditory fatigue. Some of the advantages of this approach and a brief history have been given in an earlier paper (**5**).

One of the major problems raised by partial stimulation deafness is the nature of the recovery process after stimulation. Goran de Maré (**2, 3**) first recognized that after even rather mild stimulation (at intensities previously thought to cause no post-stimulation threshold shift) there was a definite, though very brief, time before complete recovery. He found that after stimulating the ear with a 500-cps tone at a sensation level of 88 db, fatigue declined from 14.9 db at no delay to 6.16 db after delay of .7 sec. De Maré argued that this effect was neither a decline of sensation nor a persistence of vibration in the sound conduction apparatus. Bronstein and Churilova (**1**) used a stimulating tone of 2-min. duration at 94-db loudness level, and found that the recovery process was relatively delayed for the higher frequencies. Lüscher and Zwislocki (**6, 7, 8**) systematically attacked the problem of recovery from fatigue caused by short duration stimulation. The stimulating tone was always .4 sec. in duration. Under one specific fatiguing condition, namely, with 3000 cps at 80-db intensity, fatigue decreased in an exponential manner from 39 db after 20 msec. to zero db after 200 msec. The absolute amount of fatigue depends upon several factors, such as the frequency and intensity of the stimulating tone, but for the present purpose we especially note Lüscher and Zwislocki's conclusions that the recovery from fatigue is similar at all frequencies, following very nearly an exponential course irrespective of the initial absolute value of the fatigue.

The present experiment is an extension of Lüscher and Zwislocki's work, designed to explore more thoroughly the relations among fatigue, interval after stimulation, intensity of stimulation, and frequency.

Apparatus

It was desired to present to *S* a pair of tones, each controllable independently as to frequency and intensity. The first of the pair was to be the stimulating tone, the second was to be a test or probe tone, the response to which would establish the momentary auditory threshold a short while following the offset of the stimulating tone.

Two Western Electric 6B audiometers were connected through a mixing network to a pair of Permoflux PDR-8 phones, fitted with supra-aural cushions, one for *E* and one for *S*. A three-channel predetermined electronic timer made by Potter Instrument Co. was used to control the duration of output of each audiometer, and of the interval between outputs. The audiometer dials themselves, in 2.5-db steps, controlled the frequency and intensity of the two tones. The electronic timer was placed outside the sound-proof room and operated by a microswitch in *E*'s hand. The frequency of the audiometers was set with reference to a Hewlet-Packard direct-reading frequency meter, accurate within ±1 per cent.

Subjects and Procedure

Four normal-hearing *S*s were used at the frequencies 500, 3000, and 8000 cps. Duration of the first or fatiguing tone was always 400 msec., of the second or probe tone always 50 msec. The fatiguing tone was either 10-, 30-, 50-, or 70-db sensation level. The interval of silence between

tones, during which recovery proceeded, varied from 20 to 300 msec. The frequency chosen for any experimental session was set on one audiometer with the frequency meter, and the other adjusted to it by the method of aural beats.

Three thresholds were determined in each experimental session, all by the method of constant stimuli (five judgments at 2.5-db intensity steps in random order): (a) for the fatiguing tone alone; (b) for the probe tone alone; and (c) for the probe tone preceded by the fatiguing tone.

The threshold shift between (b) and (c) was taken as the amount of fatigue. Five fatigue determinations were averaged for each combination of frequency, intensity, and recovery interval.

Results and Discussion

Figure 1 shows the rise in fatigue with increasing intensity, for seven recovery periods. With the shortest recovery period of 20 msec., a fatigue effect can be found even though the stimulating tone is only about 5 db above threshold. Notice also that as the intensity of the fatiguing tone is increased, each curve in Fig. 1 exhibits a double slope. Nothing of the sort occurs in the data of Lüscher and Zwislocki (**7**, p. 433), but an even more perplexing shape appears in the data of Gardner (**4**, p. 184), and of Munson and Gardner (**9**, p. 180–181), where what these workers term a

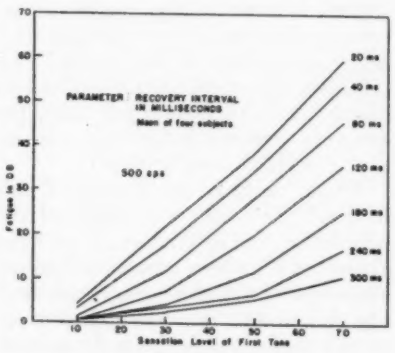

Fig. 1. Effect of intensity of stimulation upon auditory fatigue

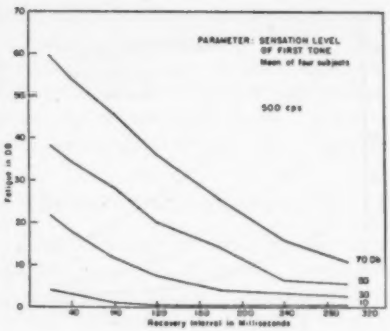

Fig. 2. Recovery from auditory fatigue at 500 cps

"saturation effect" occurs, such that as the fatiguing tone is raised from about 40- or 50- to about 80-db sensation level, relatively little additional fatigue is occasioned. The data of Munson and Gardner were gathered at recovery periods of about .1 and .2 sec., and they state that "tests at shorter delay times indicate that saturation becomes less noticeable as this factor decreases until at zero time (i.e., for simultaneous masking) it is difficult to detect or entirely absent" (**9**, p. 180). However, it is seen in Fig. 1, and the same holds true for the other two frequencies, that the double-slope feature is maintained at all recovery periods used.

It is clear that data such as in Fig. 1 may be recast in another form to show the course of decline of fatigue. This is done for the three frequencies in Fig. 2, 3, and 4. The figures are interpreted as follows: at 500 cps, when the first tone is of 70-db sensation level, fatigue has reached a level of about 59 db after 20 msec., has declined to about 36 db after 120 msec., and has reached a zero value at 360 msec. The differences between the sets of data in Fig. 2, 3, and 4 may in part be attributed to the differential effect of

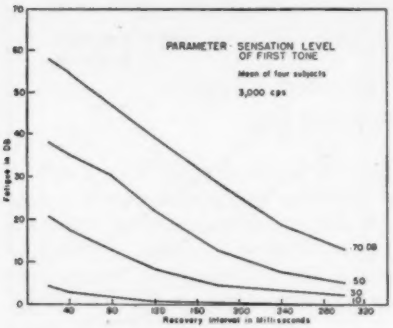

FIG. 3. Recovery from auditory fatigue at 3000 cps

frequency at constant sensation level, as described in our earlier paper, and in part perhaps to the status of the particular ear at the several frequencies.

The exact slope of the decline of fatigue is a matter of considerable interest. In general, it appears to follow a regular decline in decibels with time.

In order to check the universality of this point, and to get an idea of individual differences, the decline of fatigue for a 50-db sensation-level tone of 2000 cps was determined for eight *S*s. The average fatigue at each

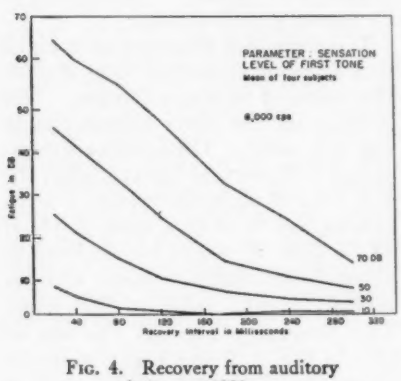

FIG. 4. Recovery from auditory fatigue at 8000 cps

recovery interval, and the standard error, are shown in Fig. 5. It is clear that the best fitting curve is a straight line for this particular sensation level, and that there are relatively slight individual differences in the course of the disappearance of fatigue. These eight *S*s all had essentially normal hearing; it remains to be determined whether the decline of fatigue takes this same form in cases with various types of deafness. There are, however, systematic deviations from this straight-line recovery, especially at the lower intensities. At 30-db sensation level, for example, Fig. 2–4 show that for all frequencies the recovery

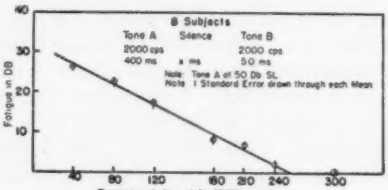

FIG. 5. Average recovery from auditory fatigue at 2000 cps

proceeds with negative acceleration even when fatigue is plotted on a log scale. Of course, even when as at higher intensities the course of recovery appears as a straight line on a log scale, it must be understood that if fatigue were to be plotted in absolute terms rather than on a log scale, the recovery process would still assume sharp negative acceleration.

What, then, is the nature of this fatigue? It is not a function of the middle ear; in our laboratory we have subjected a patient with a radical ossilectomy to the same conditions as of Fig. 5, and find that the course of recovery does not differ from the normal. Experiments on cross-conduction have ruled out a central locus. Evidently

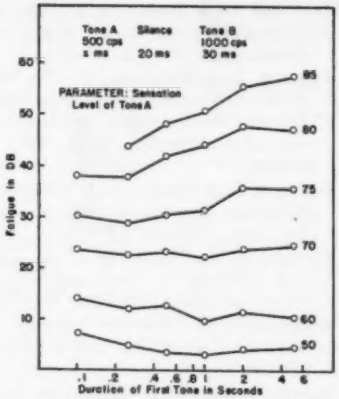

FIG. 6. Effect of duration of stimulation on auditory fatigue

the site is within the cochlea. Rosenblith (**10**) and Rosenzweig and Rosenblith (**11**) have shown recovery curves of the nerve response to clicks in the animal, both from the round window and the cortex. These curves bear some correspondences to the recovery data of this paper; but it is not clear that the fatigue process is of nervous origin. A decline in gross nerve response would occur as well if the site of the fatigue were, say, in the organ of Corti, or in the transmission mechanism between the organ of Corti and the first-order neuron.

Further information about the fatigue discussed here can be obtained from Fig. 6, which considers the effect of duration of the fatiguing tone. It is seen that the usual time-intensity relationship is here violated. Evidently the ear, though suffering from a process which reduces its threshold, is able to make an adaptation up to a point so that at moderate intensities no further fatigue will result from indefinitely prolonged stimulation. At somewhat higher levels the adaptation can be maintained though not for as long, and, of course, at extremely high intensities damage may result. But for intensities short of this extreme, fatigue must be a reversible, and in some cases noncumulative, adaptation, probably of the peripheral organ.

SUMMARY AND CONCLUSIONS

The technique of studying auditory fatigue using brief, weak tones has proved very fruitful, since the ear's return to normal is so quick that a great deal of information can be gathered in a short time. This paper studies the course of the recovery from such fatigue as a function of stimulation frequency and intensity, and of recovery interval.

To a pure tone of 500, 3000, or 8000 cps up to 70-db sensation level, acting for .4 sec., recovery is complete within little more than a third of a second. For weaker stimuli recovery is proportionately quicker. At about 50-db sensation level, the disappearance of fatigue follows roughly an exponential course; at weaker levels a negative acceleration appears. Experiments with cross-conduction and on a patient with ossicles missing rule out the central nervous system and the middle ear respectively as the locus of fatigue. An experiment on the effect of duration of stimulation shows that the ear within at least 100 msec. or less exhibits a residual threshold shift due to stimulation, but that at moderate intensities no greater threshold shifts may be occasioned with fatiguing stimulus durations of at least 10 sec. and probably indefinitely. Whether this continuously adapting mechanism is within the peripheral organ, at the transmission between the sense cells and the first nerve fibers, or is purely neural, is not yet known.

(Manuscript received May 21, 1951)

References

1. Bronstein, A. J., & Churilova, E. A. The dependence of time taken to restore the initial excitability of the hearing apparatus upon the pitch of the acting tone. *Bull. Biol. Med. exp.*, 1936, **1**, 428–430.
2. de Maré, G. Audiometrische Untersuchungen. *Acta oto-laryng., Stockh.*, 1939, Suppl. 31, 1–173.
3. de Maré, G. Fresh observations as to the so-called masking effect of the ear and its possible diagnostic significance. *Acta oto-laryng., Stockh.*, 1940, **28**, 314–316.
4. Gardner, M. B. Short duration auditory fatigue as a method of classifying hearing impairment. *J. acoust. Soc. Amer.*, 1947, **19**, 178–190.
5. Harris, J. D., Rawnsley, A., & Kelsey, P. Studies in short duration auditory fatigue: I. Frequency differences as a function of intensity. *J. exp. Psychol.*, 1951, **42**, 430–436.
6. Lüscher, E., & Zwislocki, J. Über Abklingvorgange des Ohres. *Practica oto-rhino-laryngol.*, 1946, **8**, 531–533.
7. Lüscher, E., & Zwislocki, J. The decay of sensation and the remainder of adaptation after short pure-tone impulses on the ear. *Acta oto-laryng., Stockh.*, 1947, **35**, 428–445.
8. Lüscher, E., & Zwislocki, J. Adaptation of the ear to sound stimuli. *J. acoust. Soc. Amer.*, 1949, **21**, 135–139.
9. Munson, W. A., & Gardner, M. B. Loudness patterns—a new approach. *J. acoust. Soc. Amer.*, 1950, **22**, 177–190.
10. Rosenblith, W. Auditory masking and fatigue. *J. acoust. Soc. Amer.*, 1950, **22**, 792–800.
11. Rosenzweig, M. R., & Rosenblith, W. A. Some electro-physiological correlates of the perception of successive clicks. *J. acoust. Soc. Amer.*, 1950, **22**, 878–880.

STUDIES OF DISTRIBUTED PRACTICE: V. LEARNING AND RETENTION OF CONCEPTS

LEONARD OSEAS AND BENTON J. UNDERWOOD

Northwestern University

Previous studies in this series have examined the influence of short intertrial rests on learning adjectives presented serially (**7, 8**), as paired associates (**6**), and in verbal-discrimination lists (**9**). In only serial learning was it shown that distributed trials produced faster learning than did massed trials. Theoretically, the results have been negative in that we have found no substantial support for any contemporary theories of verbal learning.

The present study was designed to extend our empirical generalizations concerning the influence of distributed practice. Serial, paired-associate, and verbal-discrimination lists emphasize what is commonly called rote learning in that the discovery of relationships is minimized and the fixation process maximized. Concept learning, on the other hand, requires that *S* discover relationships among stimuli before the response can be fixated. So far as we have been able to determine there are no published studies in which the relative influence of massed and distributed practice has been evaluated for concept learning. Experiments varying the time interval between trials in solving puzzles (**2, 3**) have shown that massing tends to be more efficient than distribution. The solution of these puzzles, like the learning of concepts, requires the discovery of relationships. However, it is questionable as to whether or not any generalization is possible from the puzzle-solving results since the intertrial interval has been inadequately explored. Massed practice, in the above-cited experiments, consisted of a few seconds between trials, while distributed practice in no case involved intervals less than 24 hr. between trials. In short, intervals which have shown facilitation in rote learning (a few seconds to 2 min.) have not been systematically investigated for tasks requiring the apprehension of relationships.

In the present study simple concepts have been learned under four different conditions of distribution, namely, 6, 15, 30, and 60 sec. between each trial.

Procedure

Materials.—Two sets of materials were used, one forming a practice list and one an experimental list. The stimuli for the experimental list consisted of three geometric forms: triangles, squares, and circles. Each form appeared in three sizes: ¾ in., ½ in., and ¼ in. in height. Hereafter these different sizes will be referred to as small, medium, and large. Finally, each form and size appeared in three shades (black, gray, white). Since all combinations of the three dimensions (form, size, shade) appeared, a total of 27 different stimuli was used. Of the three dimensions, only size and shade were relevant to the concepts to be learned; form was the irrelevant or distracting characteristic. All concepts, therefore, were determined by combinations of size and shade, and since all combinations of size and shade were used, one for each concept, there was a total of nine concepts to be learned. For example, large black square, large black circle, and large black triangle were all illustrations of a single concept (large black form). Thus the total of 27 different stimuli represented three illustrations of each of the nine concepts.

As concept indicators (responses to be made by the *S*s) single consonant letters were used. These were assigned so that obvious associations between the concept and the letter were not present. The concepts, and the letters assigned each, were as follows: large black form—R; large gray form—H; large white form—N; medium black form—J; medium gray form—Q; medium

white form—X; small black form—Z; small gray form—P; small white form—Y.

The stimulus figures were drawn in ink on white vellum tape; the concept indicators were typed on the tape. The 27 stimuli were repeated twice on the tape so that 54 figures appeared. These were divided into six groups of nine each, each group being separated from the next group by a strip of masking tape. The order of appearance of an illustration of a given concept varied for each of the groups, but each concept was illustrated once and only once in each group of nine figures. Furthermore, the same example of a given concept never appeared in two successive groups. These arrangements made it highly unlikely that the concepts could be learned by rote.

The practice-list stimuli also varied in three dimensions. These stimuli consisted of a straight inked line with small dots contiguous. The direction of the line varied three ways (vertical, horizontal, oblique); the position of the dots varied three ways (above the line, at the center of the line, below the line); and the number of dots varied from one to three. Number of dots was the irrelevant dimension, the direction of the line and the position of the dots being relevant. In other details of construction the practice list was the same as the experimental list. None of the letters used as concept indicators was the same for the two lists.

Lists were presented by means of a Hull-type memory drum at a 3:3-sec. rate, i.e., 3 sec. for the stimulus and 3 sec. for the stimulus and response together. Learning was by the anticipation method.

Subjects and conditions.—Four conditions comprised the experiment, being distinguished only by differences in the time lapse between each trial. A trial was counted as the presentation of one group of nine stimuli. Four groups of *Ss*, each group made up of 18 college students, were used. *Ss* were assigned at random to the four groups. In Group I, 6 sec. elapsed between each trial; this will be called the massed condition. For Group II the intertrial interval was 15 sec., for Group III, 30 sec., and for Group IV, 60 sec.

The *Ss* were instructed carefully on the practice day concerning the nature of a concept and differences between rote learning and concept learning were emphasized. Following these instructions the practice list was presented with *S* instructed to anticipate after the first trial. The list was presented by massed practice (6 sec. between trials) until four of the nine concepts were correctly anticipated on a single trial. From this point on learning took place by distributed practice (15 sec. between trials) until a perfect recitation was achieved. During the 15-sec. interval between trials, *S* named colors from a color board made up of rows of small patches of different colors. These were called out at the rate of 72 per min. as paced by a metronome. This same color-naming procedure was used by Groups II, III, and IV on the experimental day. After attaining the criterion on the practice list, *S* rested 10 min. and then recalled and relearned under massed conditions.

The experimental session followed the practice session by not more than three days and not less than one day. On the experimental day the experimental list was presented under the conditions specified above for each group until two successive perfect trials were attained. Each *S* returned to the laboratory 24 hr. later and relearned the experimental list by massed practice.

Results

Practice-day performance.—The mean number of trials required to learn the practice list of concepts by Groups I through IV was 13.50, 12.83, 12.39, and 12.94, with σ_M's of 1.21, 1.53, .91, and .91, respectively. The within-groups variance is much larger than the between-groups variance ($F = .16$).[1] It may be concluded that the four groups, constituted of randomly assigned groups, did not differ significantly on practice-day performance.

The original plan of the experiment called for matching the four groups on practice performance if random assignment did not result in comparable groups. Such a procedure obviously assumes that performance on the practice and experimental lists will be highly correlated. Computations of product-moment correlations have shown that this is an invalid assumption. The r's ranged from .39 for Group I to .08 for Group III. The value for all 72 *Ss* combined was .25. While we may expect the correlations

[1] With 3 and 68 *df*, an *F* of approximately 2.74 is needed for the 5% level of confidence, and 4.08 for the 1% level. On this and all subsequent analyses Bartlett's test for homogeneity of variance was made and was in no case found to be significant.

TABLE 1

LEARNING AND RETENTION SCORES FOR THE EXPERIMENTAL CONCEPTS

Interval between Trials during Learning	Learning				Retention			
	Tr. to 1 Perfect		Tr. to 2 Perfect		Recall		Tr. to Relearn	
	Mean	σ_M	Mean	σ_M	Mean	σ_M	Mean	σ_M
6 sec.	25.44	1.83	29.89	1.61	7.00	.38	2.83	.32
15 sec.	21.17	1.50	27.00	2.33	7.44	.31	2.78	.49
30 sec.	20.67	1.82	23.56	1.85	6.67	.26	3.83	.46
60 sec.	22.78	2.23	26.94	2.76	7.00	.34	2.33	.24

to be lowered somewhat because of different conditions on the two days, nevertheless, a correlation of .39 for Group I (where conditions were the same) is unsatisfactory as a basis for matching. We must conclude that the performances on these two tasks do not have high communality. Because of this, the evaluation of the results for the experimental day must rest largely on the assumption that the random assignment of Ss resulted in groups whose performance would not have differed significantly had they learned the experimental list under identical conditions.

Learning and retention of experimental list.—Data on learning and retention of the experimental list for the four conditions are shown in Table 1.

The learning scores show that on both criteria the distributed-practice conditions produced slightly faster learning than the massed condition (6 sec.). However, tests of significance for these criteria give no basis for rejecting the null hypothesis. For the criterion of one perfect trial F is 1.28, and for two perfect, 1.41. Nevertheless, the learning curves (Fig. 1) show small but consistent differences between the massed condition

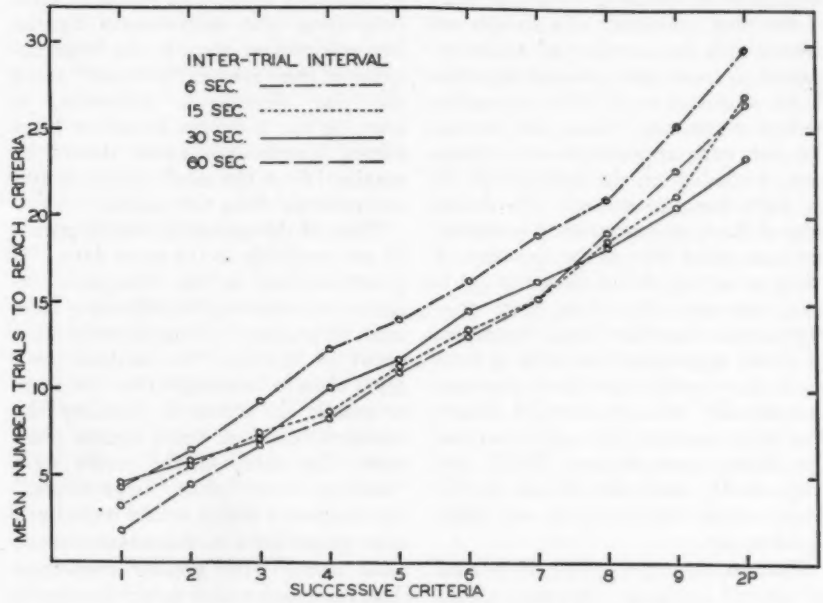

FIG. 1. Trials-to-criteria learning curves for the four conditions

and all three distributed conditions throughout the major course of learning. That these three curves of distributed practice would all fall below the curve of massed learning is highly unlikely on a chance basis if no true difference existed. It is our conclusion that distributed practice has produced a slight facilitation in the learning of simple concepts.

Cook (2) has suggested that in tasks requiring the discovery of relationships massing should facilitate early in learning and distribution late in learning. The learning curves for the acquisition of the present list of concepts give no support to this hypothesis.

The scores for recall and retention show no effect of variation in intertrial interval during learning. None of the differences attains significance. For recall F is .95, for relearning, 2.65.

Order of attainment.—An analysis was made of the order of attainment of the nine concepts. To do this we determined the number of trials required to learn each concept by each S to a criterion of two successive perfect responses. Since the results did not vary appreciably as a function of conditions, the data for all 72 Ss have been combined. To determine if shade was a source of variance, we calculated the mean number of trials to attain the three white (outline) concepts, the three gray concepts, and the three black concepts. In these determinations size is constant since each of the three sizes occurs equally often with each shade. The mean number of trials to attain the white concepts was 14.23, the gray, 14.31, and the black, 14.55. These values obviously do not differ significantly.

Size, on the other hand, is related to rate of learning. Holding shade constant, the mean number of trials to learn the concepts based on all small figures was 13.74; the medium-size concepts required 16.64 trials, and the large, 12.71. These three distributions could not have come from the same population (F = 20.51). As shown by t-tests, differences between the means for the small and medium, and between the large and medium sizes are highly significant. The t-values are 5.29 and 3.90 respectively. The difference in rate of learning the small and large concepts is not significant (t = 1.39).

These results suggest the possibility that differences in rate of attainment might be attributed to interference produced by stimulus generalization. In learning concepts of medium size, generalizing tendencies (of maximal strength for the present experiment) from both the small and large figures could interfere. But, in learning the concepts based on small figures, generalizing tendencies would be maximal from the medium-size figures, but minimal as regards the large because of their greater "distance" along the size dimension. Likewise, in learning the concepts based on large forms, interference again should be maximal from the medium-size figures but minimal from the small.

Tests of the generalization hypothesis are available in the error data. If generalization is an adequate explanatory concept, the following facts must be present: (a) more errors must occur in learning the medium concepts than in learning either the large or small; (b) errors in learning the concepts based on small figures must show that more of the errors were "medium errors" than "large errors," i.e., responses which would have been appropriate for a medium-size concept must occur with greater frequency than responses which would have been appropriate for a large figure; and (c)

errors in learning the concepts based on large figures should show that more of the errors were medium errors than small errors.

In testing these expectations we counted all errors for all *Ss* for all conditions. All three expectations are confirmed. In learning the large concepts, 1084 errors were made; in learning the medium concepts, 1440, and in learning the small concepts, 1186 errors occurred. In learning the large concepts, 406 errors were responses appropriate to the medium-size figures, and 206 errors were appropriate to the small. In learning the small concepts, 467 responses appropriate to the medium concepts were made and only 179 which were appropriate to the large. These findings strongly support the generalization hypothesis.

Discussion

The two major findings of the present study are: (a) conditions of distributed practice produce consistently faster learning than does the condition of massing; (b) concepts based on large and small figures are acquired more rapidly than those based on medium-size figures. Brief comments on the second finding will be given first.

Analysis of the errors made in learning strongly support the position that differences in order of attainment of concepts as a function of size are produced by interferences resulting from stimulus generalization. This confirms Buss's results (1). Buss demonstrated a gradient of stimulus generalization along a height dimension in concept learning. It seems clear that generalization is useful as an explanatory concept for all levels of learning, from conditioning through concept formation.

The results which show that distributed practice produced somewhat faster learning of concepts than did massed practice are completely compatible with previous findings of rote learning. The learning of serial adjectives of low intra-list similarity (7, 8), and the learning of paired-associate nonsense syllables (4) show the same small but consistent differences in favor of distribution as found in the present study. Distribution produces a more marked effect in the learning of serial adjective lists of high intra-list similarity (8), and no effect on paired adjective (6) and short verbal-discrimination lists (9).

The fact that concept learning responds to distributed practice in much the same way as rote learning does not necessarily mean, of course, that the same basic processes are required in both forms of learning. On the other hand, descriptive differences between rote learning and concept formation (as pointed out in the introduction) do not irrevocably point to fundamentally different learning processes. If further tests show that concept learning and rote learning respond in much the same way to manipulable variables, the position that we are dealing with the same basic learning processes becomes more acceptable. Such an outcome would be highly desirable from the standpoint of parsimony in theory construction. The previous studies in this series have shown that present theories are not entirely adequate to account for the effects of distributed practice. The study reported here simply extends the range of empirical facts which will need to be encompassed by a comprehensive theory.

A final comment about the retention results is needed. The present data showed that distributed practice during learning had no influence on

recall or relearning of the concepts after 24 hr. However, it is our opinion that the present study was an inadequate test of the possible relationship because retention under all conditions was quite high. This finding is in line with Reed's results (5) and it recommends that retention intervals longer than 24 hr. must be used to determine the influence of distributed practice on the retention of concepts.

Summary

To extend the empirical generalizations concerning the influence of distributed practice, simple concepts were learned with four different intertrial rests; namely, 6, 15, 30, and 60 sec. The stimuli were three different geometrical forms of three sizes and three shades. Size and shade were relevant dimensions, form irrelevant. Nine concepts were learned, each being a different combination of size and shade. Single letters served as concept indicators. Four groups of 18 *Ss* each were used, a different group for each of the above four conditions. Retention was measured 24 hr. after learning. The results show:

1. Distributed conditions (15, 30, or 60 sec. between each trial) produced faster learning than did the massed condition (6 sec.). The differences were small but consistent throughout the course of learning.

2. Retention did not vary as a function of intertrial interval during learning, being very high under all conditions.

3. Concepts based on small and large geometrical forms were learned more rapidly than those based on medium-size figures. An analysis of of the errors made during learning showed this finding to be expected on the basis of stimulus generalization.

The results produced by distributed practice in concept learning were shown to be consistent with findings of the influence of this variable on rote learning. The descriptive differences between rote learning and concept learning may not reflect fundamentally different processes.

(Manuscript received May 31, 1951)

References

1. Buss, A. H. A study of concept formation as a function of reinforcement and stimulus generalization. *J. exp. Psychol.*, 1950, **40**, 494–503.
2. Cook, T. W. Massed and distributed practice in puzzle solving. *Psychol. Rev.*, 1934, **41**, 330–355.
3. Ericksen, S. C. Variability of attack in massed and distributed practice. *J. exp. Psychol.*, 1942, **31**, 339–345.
4. Hovland, C. I. Experimental studies in rote-learning theory: VIII. Distributed practice of paired associates with varying rates of presentation. *J. exp. Psychol.*, 1949, **39**, 714–718.
5. Reed, H. B. Factors influencing the learning and retention of concepts: I. The influence of set. *J. exp. Psychol.*, 1946, **36**, 71–87.
6. Underwood, B. J. Studies of distributed practice: II. Learning and retention of paired-adjective lists with two levels of intra-list similarity. *J. exp. Psychol.*, 1951, **42**, 153–161.
7. Underwood, B. J. Studies of distributed practice: III. The influence of stage of practice in serial learning. *J. exp. Psychol.*, 1951, **42**, 291–295.
8. Underwood, B. J., & Goad, D. Studies of distributed practice: I. The influence of intra-list similarity in serial learning. *J. exp. Psychol.*, 1951, **42**, 125–134.
9. Underwood, B. J., & Viterna, R. O. Studies of distributed practice: IV. The effect of similarity and rate of presentation in verbal-discrimination learning. *J. exp. Psychol.*, 1951, **42**, 296–299.

SENSITIVITY TO APPARENT MOVEMENT IN DEPTH AS A FUNCTION OF STIMULUS DIMENSIONALITY [1]

WILLIAM M. SMITH
Princeton University

Continuous change in angular size of an object for which the usual distance-cues are not available, such as a disc seen in the dark, is ordinarily perceived not as an alteration in the size of the object but as a movement of the object in the third dimension. This phenomenon of apparent movement in depth has received very little attention in the experimental psychology of visual perception, as several recent writers have pointed out (**3, 5, 7**). As compared with the extensive literature on the phi-phenomenon, for example, studies of the phenomenon of apparent movement in the third dimension are few and far between. In addition, very little is known about this particular kind of movement regarding sensitivity functions.

One study of illusory movement in depth (**8**) showed that sensitivity to the onset of movement did not vary significantly for objects which in everyday life are frequently seen in flight or in some kind of movement. A baseball, for example, was not responded to any quicker with respect to onset of movement than were meaningless objects such as a disc. There were a few indications, however, that sensitivity in this respect might vary along a continuum of tridimensionality. Test objects with an obvious three-dimensional appearance might elicit faster reaction times than objects which appeared flat.

Such a possibility follows naturally enough from a modern extension and application of the Helmholtzian notion of *unbewusster Schluss* according to which past experience, with its wealth of associations about objects of everyday use, plays a crucial role in perception. For example, it should follow from such a view that since the everyday environment presents many three-dimensional objects in movement, the sensitivity to the movement of such objects should be greater than for objects which are flat in appearance. The main purpose of the present experiment was to determine if this difference in sensitivity would be found. The secondary purpose of the study was to determine the effect of monocular and binocular vision on sensitivity to this kind of apparent movement. Since it has been assumed by most investigators of this phenomenon that this kind of illusory movement is maximal under monocular vision and somewhat equivocal for binocular vision, it was pertinent to examine this assumption (**5**).

Apparatus and Stimulus Materials

Since the apparatus used in the present study was the same as that described in a similar experiment (**8**) the interested reader is referred there for a detailed description and diagram. The apparatus consisted of a long tablelike structure with a cubicle at one end from which *S* made observations. A superstructure on this table carried a small screen upon which was projected the desired test object from a Clason projector mounted on top of the cubicle. The projector was gear-driven by a variable-speed transmission set at a constant speed. This arrangement gave the desired size-change of the projected image. By means of a front-surface mirror and an evaporated metal film, the image on the screen was reflected to *S*'s eye(s) with the

[1] This study was supported by a grant to the Department of Psychology, Princeton University, from the Rockefeller Foundation.

apparent location of the test object directly ahead. The use of an adjustable occluding device controlled monocular and binocular viewing. Placed inside the cubicle to the right of *S* was a telegraph key which, when depressed, stopped the transmission, turned off the projector, and stopped the electric timer which measured *S*'s reaction time.

The stimulus material consisted of 35-mm. slides, one slide for each of three test objects, white on black. One figure was a Necker cube, another was a figure of the same size and outline but with internal lines radiating from the center in a spokelike fashion. The width of the lines of this figure was the same as the width of the lines making up the cube. The third figure, a solid white irregular hexagon, was of the same size and outline as the other two figures. All test figures were of the same size and shape but differed in their internal makeup in order to produce differences in apparent tridimensionality. The Necker cube appeared three dimensional, the others appeared flat, especially the spokelike figure. Prior to carrying out the experiment, independent judgments of eight individuals were taken. All judged the cube to look three dimensional, and the others more or less flat. The projected size of the test objects when initially seen by *S*, i.e., before a change in size was initiated, was 44 mm. (both dimensions). This size represented a visual angle of approximately 40 min. The rate of size-change for the test figures was 1.8 mm. per second. The brightness of the solid test figure measured by a Macbeth illuminometer was approximately one foot-lambert.

Procedure and Experimental Design

The design of the experiment required observations from six *S*s who were all college students with 20/20 vision and naive regarding the apparatus and purpose of the experiment.

A systematic latin-square design was used. The six combinations of the three test objects and the two viewing conditions were randomly ordered and rotated among the six *S*s. Each *S* made four sets of observations, each set consisting of one observation for each of the six combinations. The last three sets were replications of the first set, and only these three sets of data were used in the analysis of results. The first set was disregarded for reasons of practice.

When *S* was brought in for observations, he was blindfolded outside of the experimental room, and then led to the cubicle where he was dark-adapted for 5 min. When properly seated and adjusted in the chin- and headrest, he was given the following instructions:

> You are going to be shown different objects in front of you. Your task is to watch what I show you and at the *instant* you perceive it moving toward you, press the key under your right hand. Remember, only press the key when you definitely see what I show you moving toward you, not when you *think* it is moving, but when you *see* it moving. After you press the key, wait until I am ready for the next trial. Do not try to remove your head from the rest until I tell you to do so. Also, never try to move your head while making an observation. When I am ready, I shall show you what you are to judge and then at some later time start it moving. You should remember that what you see on any one trial will not necessarily move. In other words, on some trials everything will happen as usual, except there will be no movement. Depend only on what you actually see. Do you have any questions?

For each trial *E* would adjust the occluder for the required viewing condition and then raise the mask placed in front of *S*'s eyes between observations. The *E* would ask if *S* saw the test object, and at some variable time later *E* would engage the transmission. When *S* pressed the key, the projector light went out and the electric timer stopped. The *E* dropped the mask back in place, recorded the reaction time in .01-sec. units, reset the projector and timer and prepared for the next trial. The false trials (i.e., no size-change) were interspersed at random throughout each set of observations. The *S* was given a brief rest between each set. When the observations were completed, *S* was again blindfolded and taken outside the experimental room where he was asked a number of question pertaining to the observations.

Results

The results of the experiment consist of the reaction times to movement and supplementary verbal material obtained from each *S*. Table 1 gives the raw data for each *S* for each test object under each viewing condition. The entry in each cell is a mean of three judgments expressed in .01-sec. units. As pointed out in a previous study (**8**) the analysis of time data by statistical techniques requiring a normal distribution of data often demands a readjustment of the data by some kind of transformation. A plot

TABLE 1

MEAN REACTION TIMES OF SIX *Ss* FOR THREE TEST OBJECTS UNDER MONOCULAR (M) AND BINOCULAR (B) VIEWING CONDITIONS

S		Test Figures		
		Cube	Solid	Spoke
1	M	7.81*	5.45	7.83
	B	4.58	1.91	4.23
	Mean	6.20	3.68	6.03
2	M	12.87	13.53	11.24
	B	8.15	10.67	10.88
	Mean	10.51	12.10	11.06
3	M	4.14	4.56	3.85
	B	3.65	2.94	4.04
	Mean	3.90	3.75	3.94
4	M	8.36	8.85	9.79
	B	6.47	6.94	7.33
	Mean	7.42	7.90	8.56
5	M	8.43	14.92	14.78
	B	6.54	7.05	5.34
	Mean	7.48	10.98	10.06
6	M	10.51	5.96	5.54
	B	2.69	4.47	4.47
	Mean	6.60	5.22	5.00

* Each entry in the M and B cells is the mean of three measures in .01-sec. units.

of the present data indicated the need of such a transformation and this was done by converting to three-place log scores. Each reaction time was transformed and the three scores for each *S* under each of the six combinations of conditions were summed to give a mean log score.

Before carrying out the main analysis, it was necessary to determine whether the use of the latin-square design imposed any sequence-effects on the data. The presence of such effects would make it difficult to assess other effects unless they were known and allowance made for them. Consequently, an analysis of variance was made on a 6 × 6 latin square, each cell in the square representing the mean log time of three observations. The results of this analysis gave a mean square for sequence of .021 (5 *df*) and a mean square remainder term of .017 (20 *df*). The *F*-ratio for such results does not approach statistical significance.

After it was decided that any effect due to sequence would not disturb subsequent analysis, the transformed data were arranged in a triple classification table and an analysis of variance carried out. The results of this analysis are given in Table 2.

It can be seen from the table that none of the first order interactions is significant when tested against the triple interaction term. Main effects, however, when tested against this term give *F*-ratios significant at the .001 level for the effects of viewing conditions and *Ss*. The variance due to test objects, however, does not approach significance. This result indicates no real differences in reaction times among the test objects differing in apparent dimensionality.

The extremely large variance estimate for the effect of viewing condition, which is larger than all other sources combined, should be noted.

TABLE 2

ANALYSIS OF VARIANCE FOR TRANSFORMED REACTION TIME DATA

Source	Sum of Squares	*df*	Variance Estimate
1. Conditions (monocular and binocular)	.417	1	.417*
2. Test figures	.050	2	.025
3. *Ss*	.949	5	.190*
4. 1 × 2 interaction	.021	2	.010
5. 1 × 3 interaction	.117	5	.023
6. 2 × 3 interaction	.146	10	.015
7. Triple interaction	.179	10	.018

* Significant at the .001 level.

The difference between these two conditions consistently (one exception out of 18 comparisons) favored binocular vision.

The results given above can be supplemented somewhat by answers to questions asked of each *S* when he finished making observations. No attempt will be made to classify these answers systematically, but their general trend can be pointed out. Four of the six *S*s, for example, stated that one particular test object seemed to move more readily than the others. Of these four *S*s, three stated a preference for the cube, the fourth *S* chose the solid. However, of those *S*s preferring the cube, their reaction times did not agree with their preference. Just as the data showed no consistently faster reaction time to any particular test object, the post-experimental judgments likewise did not reveal a consistent preference for one particular object. However, all *S*s agreed that the cube appeared three-dimensional whereas the other two test objects appeared more or less flat.

With respect to the two viewing conditions, three of the six *S*s stated that they were aware of changes of this factor from trial to trial, and that the movement was always easier to detect under the binocular condition. The remaining *S*s who stated that they were not aware of this factor nevertheless complained that on some trials their vision "didn't seem quite right," or some statement to that effect. These trials were presumably the monocular ones. Regardless of whether the factor of viewing was noted or not, no *S* was aware of the illusory nature of the movement observed, nor guessed the real purpose of the experiment.

Discussion

In perceptual theory, the view that factors other than those definable in terms of present stimulation affect the final percept is not new. These "non-stimulus" factors are usually referred to some central mechanism which supplements, corrects, distorts, or "overlays" the data given in stimulation. It is at this point of the perceptual process that proponents of an empiricist view of perception usually maintain that past experience, learning, etc. manifest their effects. Thus, for example, one explanation with regard to the phenomenon of color constancy might run as follows. We see an orange; such objects have a distinctive hue which has been learned. When seen under monochromatic illumination in an otherwise normal setting, the learned response in the form of a "correction" or memory mechanism carries the perception despite the contradiction in stimulation, i.e., the orange is perceived as normal in color. The view is that in the past we have learned that a particular object has particular properties. These properties, of course, originally emerge from the characteristics of stimulation. Once these properties are learned, subsequent perception of the object brings this learning into play. In the extreme case, the learned response may contradict the data given in present stimulation, and still determine the perception. In the context of the present experiment, the view would be that we have observed the movement of three-dimensional objects countless times in the past but not objects which appear flat. Given equivalent stimulating conditions for two such kinds of objects, we might expect that the sensitivity to the onset of movement for

an object with an obvious three-dimensional appearance would be greater than for an object with a flat appearance.

The present results show, however, that the apparent tridimensionality of an object has no significant effect on the sensitivity to apparent movement in depth. These results supplement the findings reported in a similar study (**8**) where it was shown that movement sensitivity did not vary significantly with the "movement property" of an object over a variety of conditions of observation. It was true, however, that reaction times were consistently faster under *one* condition of observation for the stimulus designated as having the "property of movement," i.e., the baseball. Had a variety of conditions been investigated in the present study, it is possible, too, that some one condition might have given faster reaction times to the object appearing three-dimensional.

In a study by Lindemann (**6**) it was found that apparent dimensionality of a figure did affect the perception of movement, in this case, gamma movement, a type of apparent movement wherein an object appears to change in size with a change in illumination. Under certain conditions, this change is also accompanied by apparent movement toward *S*. This investigator reports that when a figure was apprehended as a *thing*, it tended to remain static and when apprehended as a surface, it tended to expand or contract. In addition, he reports that gamma movement was less for a figure which appeared flat.

These results for gamma movement would lead one to expect results different from those found in the present study and opposite to the results predicted by an empiricist hypothesis. They would suggest that the sensitivity to apparent movement in depth would be greatest for the test objects having a flat appearance. Present results, however, show no real difference among the test objects and hence are in disagreement with the gamma movement data and the results expected from the empiricist point of view. Nevertheless, it is possible that differences among test objects would have been evident had the movement been observed over a period of time. Under such conditions the quality of the movement might have varied systematically with the apparent dimensionality of the test object, as in the case of gamma movement. One might hypothesize that such a factor as apparent dimensionality does not affect sensitivity functions but may affect the movement experience once movement is detected and observed for a period of time.

With regard to the marked difference found between monocular and binocular vision, it was not exactly expected that a difference, if found, would favor the latter. As pointed out by Ittelson (**5**), the movement in depth observed by his *S*s was quite equivocal for the binocular condition, yet in the present study the reaction times for this condition were considerably and consistently faster than those for the monocular condition. On the basis of a perceptual theory of cues involving disparity of retinal images and allied binocular indications, it would be expected that binocular vision for this kind of movement would favor the perception of flatness and alteration in size with no movement, unless, of course, a considerable viewing distance were involved. Again the important factor may be the difference in detecting initial

change and the character of the change observed over a period of time. It is important to note that *S*s in the present experiment were set by instruction to observe movement and not size-change. Yet by the nature of the task, their discrimination was of initial change in angular size. Thus, had the present *S*s observed the phenomenon for some time after detection, the quality of the movement might have been much depressed under binocular regard, even though detected more readily under this condition. For *real* movement in depth, sensitivity is greater for binocular vision, as has been shown, for example, by Fallert (**2**). However, as pointed out in studies of visual acuity and other visual functions, the inferiority of monocular vision may result not from the use of one eye as such but from the introduction of rivalry, intensity, and other effects by the very act of closing or occluding the other eye (**1, 4**). In any event, the results of greater binocular sensitivity to this kind of apparent movement do not necessarily make it less understandable that the illusory effect is best observed under monocular regard.

Summary and Conclusions

An experiment was carried out to determine whether the sensitivity to apparent movement in depth varied with the apparent tridimensionality of objects viewed monocularly and binocularly.

An analysis of variance of the results, supplemented by statements from the *S*s, indicated the following conclusions.

1. No significant variance in reaction times was found to exist among the test objects used. Apparent tridimensionality of an object had no significant effect on the time it took *S* to respond to apparent movement in depth. This result was discussed from the standpoint of the effect of object properties on perception and an empiricistic view of perception.

2. Viewing condition (monocular and binocular) introduced significant variation in the data. The reaction times to movement were consistently faster for the binocular condition.

3. An attempt was made to interpret the present results in light of other experimental results which indicate an effect on movement perception of such a factor as apparent dimensionality. In addition, an attempt was made to incorporate the results regarding the effect of viewing conditions with other pertinent binocular-monocular data.

4. Although no significant interactions existed in the data, marked individual differences in reaction times to this kind of apparent movement were found.

(Manuscript received for priority publication October 15, 1951)

References

1. Allers, R. Über einige Unterschiede zwischen dem ein- und dem beidäugigen Sehen und über den Einfluss seelischer Momente auf einfache Leistungen des Gesichtssinnes. *S. B. Akad. Wiss. Wien*, 1935, **144**, 33–79.
2. Fallert, F. Untersuchungen über die Wahrnehmung von Bewegung in den drei Dimensionen des Raumes. *Z. Sinnesphysiol.*, 1930, **60**, 297–324.
3. Hastorf, A. H. The influence of suggestion on the relationship between stimulus size and perceived distance. *J. Psychol.*, 1950, **29**, 195–217.
4. Horowitz, M. W. An analysis of the superiority of binocular over monocular visual acuity. *J. exp. Psychol.*, 1949, **39**, 581–596.
5. Ittelson, W. H. Size as a cue to distance: radial motion. *Amer. J. Psychol.*, 1951, **64**, 188–202.

6. LINDEMANN, E. Experimentelle Untersuchungen über das Entstehen und Vergehen von Gestalten. Beiträge zur Psychologie der Gestalt. *Psychol. Forsch.*, 1922, **2**, 5–60.
7. SMITH, W. M. Past experience and the perception of visual size. *Amer. J. Psychol.*, 1952, **65**, (In press).
8. SMITH, W. M. The sensitivity to apparent movement in depth as a function of "property of movement." *J. exp. Psychol.*, 1951, **42**, 143–152.

THE ROLE OF REPETITION RATE AND INTER-STIMULUS INTERVAL IN CONTEXT EFFECTS [1]

J. M. DOUGHTY
Franklin and Marshall College

That psychophysical judgments are seriously influenced by the context in which they are made is a well-established fact. That the constant error (CE) can be moved about at will even with keen and experienced *O*s by manipulating the context is known (**1, 4**). Current interpretations of these context effects are couched in terms of interaction effects among the stimuli making up a particular experimental series (**2, 3**). That is, it is thought that stimulation that has previously occurred in some way interacts with subsequent stimulation so that *O* seems to be comparing the present comparison tone not with the actual standard but with some other standard that approximates the mean of the series. If this interpretation is correct then it is to be expected that context effects will vary as the temporal relations between the stimuli of the series are changed. For example, a previous high pitch will contrast less sharply with a subsequent low pitch if the time interval is lengthened considerably. The two ways in which the temporal separation of stimuli in the psychophysical methods may be accomplished is by changing the rate at which the stimuli are presented and altering the separation between the standard and comparison stimuli. One may hold repetition rate constant and increase the separation between the paired standard and comparison, or one may hold the separation between standard and comparison constant and change the repetition rate.

The purpose of this experiment was to explore the role of the temporal separation of stimuli on context effects. To this end repetition rate and inter-stimulus interval (the separation between standard and comparison stimuli) were varied in the methods of constant stimuli and average error. CE's and CL's for pitch were determined in a balanced and unbalanced context to see if they varied in any systematic way with repetition rate and inter-stimulus interval.

As a point of comparison CE's for pitch were also determined by the method of single stimuli. It was expected that the maximum effect of the unbalanced context would be displayed in this method since the standard stimulus is sounded only once at the beginning of a run. Thus, the context within which *O* makes his judgments of pitch is composed almost exclusively of the comparison stimuli. The modifying influence of the standard tone (repeated for each judgment in the method of constant stimuli) is absent.

Description of the Experiment

Apparatus

The apparatus used in this experiment was identical to that described in detail in a previous paper (**1**).

Procedure

Method of constant stimuli.—The *O* was seated in the sound-deadened room and asked to adjust the headphone so that it fitted comfort-

[1] This research was carried out under terms of a contract between Special Devices Center, Office of Naval Research, and The Johns Hopkins University, Contract N5-ori-166, Task Order I. This is Report No. 166-I-135, Project Designation No. NR-784-001 under that contract.

ably. He was instructed to use the same ear at all times. He was told to listen to the standard stimulus and to judge whether the comparison tone was higher or lower in pitch. He was told to record this judgment in the appropriate place on the tally sheet.

In a single run *O* made 70 such judgments; ten judgments per point at seven points. In all, there were five runs for each *O* for each set of conditions, making a total of 50 judgments per point per *O*.

Method of average error.—The procedure was the same as that for the method of constant stimuli except, of course, that *O* adjusted the variable stimulus until the two tones were matched in pitch. One judgment was made at a time and a new set of conditions was set in before *O* was asked to make another judgment. Between judgments he removed the earphones. Sets of about 12 such judgments were regularly alternated with the runs under the method of constant stimuli. Each *O* made a total of 30 judgments per point.

The *E* used the fine vernier adjustment on the beat-frequency oscillator in combination with the oscillograph in determining the *O*'s setting for his pitch matches. These were recorded in terms of the frequency to match the standard.

Experimental Variables

Degree of asymmetry: Method of constant stimuli.—This refers to the degree of skewness in the distribution of the comparison stimuli about the standard. There were two degrees of asymmetry defined in terms of the number of comparison stimuli higher than, equal to, and lower than the standard stimulus. Zero asymmetry had three higher, one equal, and three lower; +2 asymmetry had five higher, one equal, and one lower.

Degree of asymmetry: Method of average error.—For this method, there were two degrees of asymmetry in the stimulus context, defined as follows: (0)—The *O* was started at a high or low point and was instructed to bracket the standard as symmetrically as possible in making the match; (+)—the *O* was started at a point higher than the standard and was instructed to change the variable in one direction only until the match was made.

Comparison stimulus durations.—Comparison stimulus durations of 50 msec. and 25 msec. were used to provide two levels of difficulty in making judgments.

Repetition rates.—Repetition rates of 3, 5, 10, and 20 per minute were used.

Inter-stimulus intervals.—Temporal separations between standard and comparison stimuli of 0, .5, 2.0, 4.0, and 8.0 secs. were used.

Foreknowledge.—In this condition, *O* was told what was the distribution of comparison stimuli with respect to the standard stimulus; e.g., *O* was told that there were five frequencies higher than the standard for every one equal to or lower than the standard.

Experimental constants.—For all judgments the following were constant: (a) The duration of the standard tone at 1000 msec. (b) The frequency of the standard tone at 1000 cps. (c) The intensity of the standard tone at 80 db re .0002 dyne/cm.^{2} for two *Os* and 60 db re .0002 dyne/cm.^{2} for one *O*. (d) The mode of listening as monaural.

Training procedure.—Two *Os* were used throughout the experiment. They were both males about 27 years of age. A third *O*, female and in her early twenties, was used in about half the experiment.

Each *O* was given a period of training in making pitch judgments. At the same time they were gaining familiarity with the procedures for the experimental methods we used. This included experience with both comparison tone durations and with varying degrees of asymmetry. This period ended, and the actual experiment began for each *O* when he had reached the point of maximum discrimination, as defined by no further improvement.

During the training the *Os* were repeatedly instructed not to be concerned with the distribution of their judgments (the number of "highs" or "lows," for instance) but to make all their judgments only with respect to the relation between each standard-comparison pair as it came along.

In the training period for the method of average error, and throughout the experiment, the *Os* were told that their primary task was to make the equation in pitch. This task was to take precedence over moving the variable stimulus in one direction only when they were judging under the conditions of asymmetry outlined above for this method.

Analysis of data.—For the method of average error data, CE's were computed from the mean of the distribution of judgments for each of the *Os*. The DL was taken as the PE of the distribution of each *O*'s judgments. An average of the individual CE's was computed for the presentation of the results in graphical form.

The least squares solution for the phi-gamma function, with unweighted values, was used to determine the CE's and DL's from the responses of the *Os* in the method of constant stimuli. The CE was taken as the 50% point and the DL as one-half the interval between the 25% and 75% points. Each *O*'s data were analyzed in this way and an average computed for presentation in graphs.

Results

Effect of Repetition Rate

Figures 1 and 2 show the magnitude of the influence of an asymmetrical context on CE's for pitch as a function of the rate of presentation of the stimulus pairs. Figure 1 shows this relation for the methods of constant stimuli (solid line) and single stimuli (broken line) and for the method of constant stimuli with foreknowledge of the asymmetry in the context (dotted line). Figure 2 shows this relation for the method of average error.

The points on Fig. 1 are based on the judgments of two *O*s and those in Fig. 2 on three *O*s. They were obtained by subtracting the CE in cycles per second at zero asymmetry from the CE at +2 asymmetry.

In Fig. 1 some indication of inter-observer variability is shown by the vertical lines passing through the points on the graph. The terminal points of these vertical lines represent the results obtained from the individual *O*s. For example, one *O*'s results are indicated by the small horizontal line drawn through one of the terminal points of the vertical line. In Fig. 2 the size of the *SD* of the *O*s' judgments is indicated by the vertical lines.

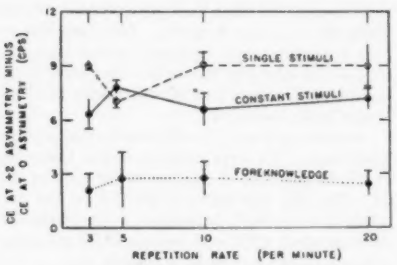

Fig. 1. The magnitude of the influence of an asymmetrical context (+2) on CE's for pitch as a function of repetition rate. The CE's are for a comparison tone of 50 msec. and an interval of 500 msec. between standard and comparison. Each curve is for a different method as indicated. The points are based on the judgments of two *O*s, whose individual results are shown by the terminal points of the vertical lines.

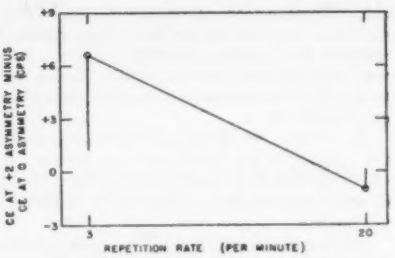

Fig. 2. The magnitude of the influence of an asymmetrical context (+2) on CE's for pitch determined by the method of average error, as a function of repetition rate. The CE's are for a comparison tone of 50 msec. and an interval of zero second between standard and comparison. The points are based on the judgments of three *O*s. The size of the *SD* of the *O*s' judgments is indicated by the vertical lines.

From these figures several conclusions may be drawn:

1. The CE for pitch as determined by the method of constant stimuli is profoundly affected by asymmetry in the context at all rates of presentation. This is true when the stimuli in the context are presented very slowly (3 per min.) as well as when they are presented very rapidly (20 per min.).

2. The effect of asymmetry on the CE for pitch appears to be as great in the method of constant stimuli as in the method of single stimuli. The points on the dotted line are not reliably different from those on the solid line in Fig. 1.

3. The effect of the asymmetry in the context on the CE for pitch is constant throughout the range of repetition rates used in the method of constant stimuli and method of single stimuli. None of the points on the individual curves in Fig. 1 is reliably different from any other.

4. In the method of average error, the effect of the asymmetry in the context is less at the fast repetition rate (20 per min.) than at the slow repetition rate (3 per min).

5. There is a significant effect of the asymmetry in the context on the CE for pitch even when there is foreknowledge of the unbalanced nature of the context. The magnitude of the effect, however, is significantly less with foreknowledge than without.

Effect of Inter-stimulus Interval

Figure 3 shows how the effect of an asymmetrical context on the CE for pitch varies with the interval between standard and comparison stimulus (inter-stimulus interval). The solid line shows the results obtained with the method of constant stimuli at two repetition rates (3 and 5 per min.) and the broken line shows the results obtained with the method of average error at a repetition rate of 3 per min.

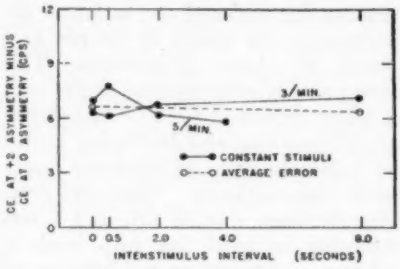

FIG. 3. The magnitude of the influence of an asymmetrical context on CE's for pitch as a function of inter-stimulus interval. This relation is shown at two repetition rates for the method of constant stimuli (3 and 5 per min. as indicated in the solid curves) and at one repetition rate for the method of average error (3 per min.). The comparison tone duration was 50 msec. in all cases. The points on the solid line are based on the judgments of two *Os*; those on the broken line on the judgments of three *Os*. The open square represents the data from the method of single stimuli obtained at a repetition rate of 3 per min.

The solid curve (method of constant stimuli) is based on the judgments of two *Os*. The broken curve (method of average error) is based on the judgments of three *Os*.

The conclusions that can be drawn from the data as displayed in Fig. 3 are similar for both methods:

1. The effect of the asymmetrical context is significant at all inter-stimulus intervals.

2. There is no systematic change in the influence of the asymmetrical context with changes in the inter-stimulus interval. The influence of context appears to be constant throughout the range of intervals used in this experiment.

Method of single stimuli.—The greatest separation between standard and comparison stimuli in this experiment is actually to be found in the method of single stimuli. For this reason, data obtained by this method have been entered in Fig. 3 in the form of the single point indicated by the open square. This point was extracted from Fig. 1 showing the shift in the CE resulting from the asymmetrical context as it occurs at a repetition rate of 3 per min.

The point represented by the open square lies above the other curves in the graph but is not reliably different from them. This result tends to confirm the conclusion that the effect of context is a constant one for all degrees of separation between standard and comparison stimuli.

DISCUSSION

As was pointed out in the introductory section of this paper, changing the temporal relations between the stimuli in the series should alter the supposed interaction between them and, consequently, the effect of con-

text. The fact that the effect of context on the CE for pitch is constant throughout the range of repetition rates and inter-stimulus intervals used in this experiment argues against the view that context effects are explicable solely in terms of interaction effects among the stimuli.

On the other hand, such results as these are consistent with the view that the *O* has a tendency to split his judgments 50–50 between the two alternative responses and that this tendency can account for context effects. With such a tendency remaining constant, it is to be expected that the effect of asymmetry in the context should be a constant one.

Do context effects disappear when such a tendency is removed? Data relevant to this question are provided by the series of judgments made at the end of the experiment with foreknowledge of the asymmetry. The results in Fig. 1, dotted line, show that even with this foreknowledge there still remains some effect of context. Is this residual effect due to interaction effects? An affirmative answer cannot be given to this question because it is conceivable that the tendency to split the judgments 50–50 has not been eliminated by foreknowledge. The *O* knows that there are some stimuli lower in frequency than the standard. He has been shown that. When he is in doubt on difficult discriminations and when he has already made a great number of "higher" judgments, he may still have a tendency to call the difficult ones "lower" more often than he would in a balanced context.

Two other facts should be borne in mind in considering the interpretation of context effects: (a) There appears to be a reliable trend toward increasing DL's with decreasing repetition rate; (b) The DL was significantly larger for the inter-stimulus interval of zero second than for any of the other intervals used. (The trend with increasing inter-stimulus intervals beyond zero second was unreliable.) Both of these facts are hard to reconcile with the view that context effects are explicable solely in terms of the tendency on the part of the *O* to split his judgments 50–50. As was pointed out in a previous paper (1) such tendency should be greater when the discriminations are more difficult. The larger DL's indicate a greater difficulty in making pitch discriminations. There should be, then, a direct relation between the magnitude of the effect of the context on the CE and the size of the DL. This, of course, is not the case.

On the other hand, the fact that there appears to be a reliable trend toward increasing DL's with decreasing repetition rate is more in line with the view that the influence of context is due to interaction effects among the stimuli in the series. The *O* responds as though a new standard, approximating the mean of the series, is developed in the course of the run as a result of these interactions. This new point of reference has to act over a longer period of time for the slower repetition rates, and decreasing acuity of judgment is to be expected under such circumstances. How the relation between size of DL and inter-stimulus interval fits into such a scheme is difficult to say because of the unreliability of the results on this point.

In the light of this discussion, no conclusions can be drawn concerning the interpretation of these context effects other than that neither one of the two views expressed here is adequate to account for all the results obtained.

Method of average error.—The data displayed in Fig. 2 show that context

is less effective when the repetition rate is fast than when it is slow. It has been pointed out (1) that the apparent context effects obtained in the method of average error may be explicable in terms of an adjustment tendency. The *O* tends to move the variable stimulus up to the edge of the interval of uncertainty and to stop there when he is approaching the standard from one direction (the condition of asymmetry in this study). This factor plus the tedium of making pitch matchings when the repetition rate is very slow (3 per min.) may result in a greater degree of undershooting at the slow rate than at the fast rate (where there is, presumably, less tedium). This would account for the results obtained.

Summary and Conclusions

This experiment was designed to explore the variables of repetition rate and inter-stimulus interval in relation to the effect of context on judgments of pitch. The difference in the CE obtained in a balanced context and the CE obtained in an unbalanced context was determined under several rates of presentation (from 3 to 20 per min.) and several time intervals between standard and comparison stimuli (ranging from zero to 8.0 sec.). Under these conditions DL's were also computed. These variables were explored in the method of average error, method of single stimuli, and method of constant stimuli. The following conclusions were reached:

1. The effect of context is constant over the range of repetition rates and inter-stimulus intervals used in this experiment in both the methods of constant stimuli and single stimuli.

2. The apparent decrease in the effect of context at faster repetition rates in the method of average error may be interpretable in terms of an adjustment tendency on the part of *O* rather than strictly in terms of context effects.

(Manuscript received for priority publication October 15, 1951)

References

1. Doughty, J. M. The effect of psychophysical method and context on pitch and loudness functions. *J. exp. Psychol.*, 1949, **39**, 720–745.
2. Helson, H. Adaptation level as frame of reference for prediction of psychophysical data. *Amer. J. Psychol.*, 1947, **60**, 1–29.
3. Koester, T., & Schoenfeld, W. N. The effect of context upon judgments of pitch differences. *J. exp. Psychol.*, 1946, **36**, 417–430.
4. Rose, G. L., & Harris, J. D. Suggestions for constructing tests of pitch discrimination. Progress Report No. 2. Bureau of Medicine and Surgery, Research Project No. NM 003 026, 10 February, 1948.

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The issues listed below are available for sale

Year	Volume	Available Numbers						Price per Number	Price per Volume
1916	1	1	2	3	4	5	6	$1.25	$7.00
1917	2	1	2	–	4	5	–	$1.25	$5.00
1920	3	–	2	–	4	5	–	$1.25	$3.75
1921	4	1	–	3	4	5	6	$1.25	$6.25
1922	5	–	–	3	4	5	–	$1.25	$3.75
1923	6	–	–	–	–	–	–	——	——
1924	7	1	2	–	–	5	–	$1.25	$3.75
1925	8	1	2	3	4	–	6	$1.25	$6.25
1926	9	–	2	3	–	5	6	$1.25	$5.00
1927	10	–	–	3	4	5	6	$1.25	$5.00
1928	11	1	2	3	4	5	6	$1.25	$7.00
1929	12	1	2	3	4	5	6	$1.25	$7.00
1930	13	1	2	3	4	5	6	$1.25	$7.00
1931	14	1	2	3	4	5	6	$1.25	$7.00
1932	15	1	2	3	4	5	6	$1.25	$7.00
1933	16	1	2	3	4	5	6	$1.25	$7.00
1934	17	1	2	3	4	5	6	$1.25	$7.00
1935	18	1	2	3	4	5	6	$1.25	$7.00
1936	19	1	2	3	4	5	6	$1.25	$7.00
1937	20	1	2	3	4	5	6	$1.25	$7.00
	21	1	2	3	4	5	6	$1.25	$7.00
1938	22	1	2	3	4	5	6	$1.25	$7.00
	23	1	2	3	4	5	6	$1.25	$7.00
1939	24	1	2	3	4	5	6	$1.25	$7.00
	25	1	2	3	4	5	6	$1.25	$7.00
1940	26	1	2	3	4	5	6	$1.25	$7.00
	27	1	2	3	4	5	–	$1.25	$6.25
1941	28	1	2	3	4	5	6	$1.25	$7.00
	29	1	2	3	4	5	6	$1.25	$7.00
1942	30	1	2	3	4	5	6	$1.25	$7.00
	31	1	2	3	4	5	6	$1.25	$7.00
1943	32	1	2	3	4	5	6	$1.25	$7.00
	33	1	2	3	4	5	6	$1.25	$7.00
1944	34	1	2	3	4	5	6	$1.25	$7.00
1945	35	1	2	3	4	5	6	$1.25	$7.00
1946	36	1	2	3	4	5	6	$1.25	$7.00
1947	37	1	2	3	4	5	6	$1.25	$7.00
1948	38	1	2	3	4	5	6	$1.25	$7.00
1949	39	1	2	3	4	5	6	$1.25	$7.00
1950	40	1	2	3	4	5	6	$1.25	$7.00
1951	41	1	2	3	4	5	6	$1.25	$7.00
	42	1	2	3	4	5	6	$1.25	$7.00
1952	43	By subscription, $14.00 (2 vols.)						$1.25	——
	44								
		List price, Volumes 1 through 42							$269.00
		30% discount							80.70
		Net price, Volumes 1 through 42							$188.30

Information about the Journal of Experimental Psychology: as noted in the table, the journal was not published in 1918 and 1919; from 1937 to 1943 it was published at the rate of two volumes per year; from 1944 to 1950, in one volume per year. Beginning in 1951 there are two volumes a year.

Information about prices: when all numbers are available, the price per volume throughout the series is $7.00. For incomplete volumes, the price is $1.25 for each available number. For foreign postage, $.25 per volume should be added. The American Psychological Association gives the following discounts on orders for any one journal:

10% on orders of $ 50.00 and over
20% on orders of $100.00 and over
30% on orders of $150.00 and over

Address orders and subscriptions to

AMERICAN PSYCHOLOGICAL ASSOCIATION

1515 Massachusetts Avenue N.W. Washington 5, D. C.

Classics Among
PSYCHOLOGICAL MONOGRAPHS

Thorndike, E. L. The Mental Life of the Monkey. 1899, #15. *$.50*

Carr, Harvey. Visual Illusion of Movement During Eye Closure. 1905, #31. *$1.25*

Watson, John B. Kinaesthetic and Organic Sensations: Their Role in the Reactions of the White Rat to the Maze. 1907, #33, *$1.00*

Shepherd, William T. Some Mental Processes of the Rhesus Monkey. 1910, #52. *$.75*

Franz, Shepard Ivory and Lafora, Gonzalo R. On the Functions of the Cerebrum: The Occipital Lobes. 1911, #56. *$1.25*

Fernberger, Samuel W. On the Relation of the Methods of Just Perceptible Differences and Constant Stimuli. 1912, #61. *$1.00*

Boring, Edwin G. Learning in Dementia Praecox. 1913, #63. *$1.00*

Langfeld, Herbert S. On the Psychophysiology of a Prolonged Fast. 1914, #71. *$.75*

Franz, Shepard Ivory. I. Symptomological Differences Associated with Similar Cerebral Lesions in the Insane. II. Variations in Distribution of the Motor Centers. 1915, #81. *$1.50*

Peckstein, Louis Augustus. Whole versus Part Methods in Motor Learning. 1917, #99. *$.75*

Kjerstad, Conrad L. The Form of the Learning Curves for Memory. 1919, #116. *$1.35*

Tolman, Edward C. Retroactive Inhibition as Affected by Conditions of Learning. 1918, #107. *$.75*

MANY OF THE EARLY MONOGRAPHS ARE OUT-OF-PRINT.
ONLY A LIMITED QUANTITY OF THE ABOVE
NUMBERS ARE AVAILABLE.

AMERICAN PSYCHOLOGICAL ASSOCIATION

1515 Massachusetts Ave. N.W., Washington 5, D. C.